USING TEXT MINING TO CHARACTERIZE ONLINE DISCUSSION FACILITATION

Norma Ming
Institute for Teaching and Learning
Nexus Research and Policy Center

Eric Baumer
Information Science Department
Cornell University

ABSTRACT

Facilitating class discussions effectively is a critical yet challenging component of instruction, particularly in online environments where student and faculty interaction is limited. Our goals in this research were to identify facilitation strategies that encourage productive discussion, and to explore text mining techniques that can help discover meaningful patterns in the discussions more efficiently at scale. Based on a close reading of selected discussion threads from online undergraduate science classes, we observed a variety of facilitation strategies associated with discussion quality. These observations informed our selection of a larger dataset of discussion threads to analyze via text mining techniques. Using latent semantic analysis to produce topic models of the content of the discussions, we constructed visualizations of the topical and temporal development of those discussions among students and faculty. These visualizations revealed patterns that appeared to correspond with specific facilitation styles and with the extent to which discussions remained focused on particular topics. From a case study focusing on six of these discussions, we documented distinct patterns in the types of facilitation strategies employed and the character of the discussions that followed. In our conclusion, we discuss potential applications of these analytical techniques for helping students, faculty, and faculty developers become more aware of their participation and influence in online discussions, thereby improving their value as a learning environment.

KEYWORDS

Computer-mediated communication, asynchronous learning networks, information and communication technologies, online discussion, text mining, latent semantic analysis, discussion facilitation, automated support

I. INTRODUCTION AND HISTORICAL BACKGROUND

Previous research on how students learn through discourse has examined the value of constructing and evaluating arguments as a powerful mechanism for learning 1, 2, 3, 4. In addition to eliciting explanations 5, 6, argumentation demands analyzing the quality of evidence provided 7 as well as the correspondence between evidence and claim. Among science educators, there is increasing interest in teaching the authentic practice of science through inquiry and argumentation 8. Studying discourse and argumentation in online science classes thus offers a particularly rich opportunity for characterizing students' learning and faculty's facilitation of this learning.

How a teacher may facilitate such discussion is itself rich with possibilities. Effective teachers employ such strategies as assessing and probing students' thinking, checking for clarity of communication, acknowledging and validating ideas, and intervening to guide attention or create opportunities to learn 9. Important strategies when facilitating online discussions include motivating participation, maintaining social presence, asking probing questions, dealing with aggressive or domineering behaviors, encouraging

equitable communication, attending sensitively to differences in class / gender / status, and providing closure in discussions 10. Other valuable facilitation techniques include modulating the voice and tone of participation, guiding the discussion's direction, and encouraging students to make connections, as well as incorporating questions that probe for deeper meaning, clarify vocabulary, explore assumptions and rationale, elucidate cause and effect, and consider implications for action 11. Additional research points to the importance of providing pedagogic and affective support such as encouraging critique and divergence, fading support appropriately, and suggesting activities to generate debate 12. Useful techniques for supporting contentful processing include helping students to identify areas of agreement and disagreement and to reach consensus and shared understanding 13. Structuring the discussion with detailed guidelines and rubrics for assessing participation was found to increase interaction and discourse 14, but giving rudimentary participation credit may induce students to engage in the minimum posting behavior necessary to earn points 15.

While there is currently no substitute for reading individual discussions to understand how they unfold, such techniques for analyzing verbal data are time-consuming and do not scale readily 1617, 18. The large quantity of discussion forum data available from online classes motivates the value of developing analytical techniques capable of processing and interpreting text data on a very large scale. Text mining may help detect and visualize patterns across many classes and instructors more efficiently, highlighting specific exchanges for subsequent close textual analysis. Such techniques have been developed in recent years for analyzing online discussions (*e.g.*, 19, 20, 21, 22) or transcripts of live classes (*e.g.*, 23). Less work has focused specifically on applying such techniques to online classrooms 3. Potential applications of this research range from tools to help faculty in facilitating class discussions, to techniques for students to review past discussions, to methods for faculty trainers to review massive quantities of discussion data quickly and easily. Here, our primary goal is to analyze discussion data to understand facilitator impact on discussions and provide feedback to improve facilitation, via computational tools that offer novel perspectives and insights. This paper presents one set of analytic techniques and demonstrates how they can be applied toward this goal.

II. METHODS

The overall approach in the research here was to apply multiple analytical techniques to selected discussion threads from a small number of classes exploring related discussion questions, in order to verify the correspondence between these techniques. For each thread, this included calculating some basic quantitative metrics on post quantity, frequency, and timing; performing text mining to create topic models and visualizations of the post relationships; and qualitative analysis of the content of discussion and participants' discourse patterns. While we did not seek to create quantitative measures of discussion quality, we were particularly interested in the following factors: consistency of focus on the original topic, depth and sophistication of analysis of relevant concepts, and evidence of participants reading and responding to each other's posts thoughtfully.

We unfortunately did not have uniform assessment data on students' knowledge of the specific concepts being discussed. Grading standards and practices varied across instructors, with final grades including many other performance and participation measures and no baseline metrics available. Assignments specific to the topic of interest also varied: Some instructors assigned individual work while others assigned group projects, and one gave quizzes while the rest requested essays. These constraints prevented measuring and comparing individual students' understanding of the specific concepts from the discussions.

A. Context

The discussion threads that we studied came from undergraduate science courses in the online degree-granting program of a large, market-based, private university. These online courses all include a

discussion forum in which students are expected to respond to questions posted by their instructor, typically with two substantive responses to each of two discussion questions per week as a minimum for earning participation credit. The default expectation for minimum initial response length is 200-300 words, although this may vary in some cases. The total course length is five weeks, with discussion questions usually being required every week. The number of active students in a given thread may range from 3 to ~25, dependent on program requirements, the timing of the discussion during the class, and course completion rates.

While the course design guide that accompanies every course includes a collection of suggested discussion questions, instructors are free to modify these questions or construct their own to give their students. Faculty are expected to participate in these discussion threads by posting at least one substantive message on five of seven days for each online week, behavior which is subject to periodic and intermittent monitoring by other faculty and administrators. These reviewers also look for active, on-topic engagement in the discussion forum; demonstration of content expertise through theoretical and/or practical examples; follow-up questions encouraging greater student participation; and timely responses to student questions. In their mandatory training and ongoing professional development, faculty receive further guidance on specific discussion facilitation strategies. These guidelines encourage faculty to keep the discussions focused, provide encouragement, give precise feedback, and ask open-ended questions. Additional recommendations describe strategies such as acknowledging individual contributions, adding new perspectives, sharing experiences, connecting concepts to related course materials and to students' own experiences, suggesting alternative solutions, disagreeing constructively, and asking probing questions. The expectations for faculty's facilitation techniques thus incorporate a range of behaviors identified in the literature as supporting effective discussion.

B. Selection of Discussion Thread Data

In previous close readings of discussions from online introductory science classes, we had observed some particularly interesting facilitation styles in the beginning biology classes, in the discussions about evolutionary theory 24. While these analyses included discussions from a broader range of science disciplines and topics, the discussions of evolutionary theory showed more variability in facilitation styles and participation behaviors, perhaps due to having more classes, larger enrollments, and longer discussions. Here, we wanted to investigate similar discussions further using text mining to see if some detectable patterns might emerge. In selecting threads for this analysis, we sought to identify a set of discussion questions that included or were as close as possible to the questions from the previous discussions we had studied. We also wanted to include the two instructors whose discussions we had previously noted as demonstrating interesting facilitation strategies, as well as other instructors with potentially different facilitation styles. One instructor asked students to discuss the flaws in Lamarck's theory of evolution, while the other instructor asked students to discuss the mechanisms of natural selection. Narrowing our search to satisfy these constraints yielded two separate sets of discussion threads, one about Jean-Baptiste Lamarck's evolutionary theories, and one about the role of natural selection in evolution.

1. Lamarck Threads

These threads were selected by searching the archived introductory biology classes for all discussion questions containing the word "Lamarck." The two questions returned from the search were:

> *Explain what is wrong with Lamarck's notion of evolution. Be aware that I am a bit of a Lamarckian at heart and believe that viruses are a mechanism for evolution!*
>
> *Both Lamarck and Darwin understood the importance of inheritance to species evolution. However, there is a subtle but critical difference between the theory proposed by Jean Baptiste Lamarck in the early 1800's and that proposed by Charles Darwin some 50 years later. In your own words, describe the differences in their theories.*

The resulting dataset consisted of seven discussion threads, which contained a total of 437 posts (mean 62.4, std. dev. 29.9, min 29, max 111) by two instructors and 108 students. All the posts from these threads were treated as a single corpus from which a topic model was built using LSA.

2. Natural Selection Threads

Searching the database for all discussion questions containing the phrase "natural selection" produced 37 threads. Restricting these results to questions that focused more specifically on mechanisms of evolution yielded seven discussion questions, three being very close variants of each other:

> *What is the role of natural selection in the theory of evolution?*
>
> *What is the role of natural selection in the mechanisms of evolution? Provide an example of how this process works.*
>
> *Please describe extensively the role of natural selection in the theory of evolution.*

Due to the difficulty in tracing the development of multiple topics within the same thread, we excluded threads where the instructor provided a choice of questions to answer. We also excluded optional threads in which the instructor did not require all students to participate.

The resulting dataset consisted of six discussion threads from the classes of three instructors about the above questions, for a total of 570 posts by three instructors and 83 students. These threads were treated as a single corpus for the topic model.

C. Calculating Quantitative Metrics on Discussion Participation

In order to capture the amount of activity and characterize participation patterns in the discussions, we calculated a variety of metrics on posting behavior for each discussion thread. These included the number of active students (counted as the number of unique student IDs that appeared at least once in the discussion), the total number of posts in the thread contributed by the instructor or by the students, the total number of words posted by the instructor and by the students, and the average word count per post (for the instructor and for the students). We also calculated the average number of words and posts per student in each thread to assess relative student activity. All of these word counts excluded signature files and text quoted from previous posts.

We also constructed two visualizations for each thread to depict the posting activity by participant and over time. One visualization represented each post as a separate point, graphing word count *vs.* time (days elapsed since initial discussion question), with students' posts in one color and instructor's posts in another color. This graph portrayed both the changes in activity in a thread over time and the relative contributions by students and instructor. The second visualization displayed the time interval between first and last post for each participant separately. This graph revealed how long each participant remained active in the thread.

D. Implementing Text Mining Techniques

The implementation presented here includes three main components:

1. a topic model, used to analyze the topics in discussion threads;
2. a projection of the post-by-post topic analysis into two dimensions; and
3. a topic space visualization, which graphs the two-dimensional projection of posts to show conceptual distance between posts.

The next section describes the organization of the data being analyzed, followed by a description of each of these components.

E. Structure of Data

The discussion threads analyzed here consisted of an initial question posed by the instructor, followed by responses by the students and usually the instructor, in threaded format. Since posts were stored as HTML, we used a variety of methods to remove HTML markup and convert HTML entities to text. We also took advantage of the HTML markup to remove automatically attached signatures and text quoted from previous posts. While not all students or instructors used the default formatting for signatures or quoting previous posts, these methods still enabled a significant reduction of noise in the data.

F. Building a Topic Model

The work reported here explores the use of computational methods for analyzing the development of topics in a discussion as one means for measuring its coherence and content. Specifically, we employ a technique called latent semantic analysis (LSA) 25, 26, 27 to perform a topic analysis of class discussion threads and reveal which topics are and are not being discussed.

Topic modeling represents a topic as a weighted set of related terms 28, 29. Some words are highly likely to appear in documents about a given topic, while other words are highly unlikely to appear. For example, in documents about pet animals, a topic about cats might be associated with such terms as “cat,” “purr,” “meow,” “mouse,” or “kitten,” while a topic about dogs might be associated with such terms as “dog,” “bark,” “woof,” “bone,” or “puppy,” and terms such as “fur,” “collar,” or “tail” might be associated with both topics. A single document that discusses both cats and dogs would then be represented as a weighted mixture of these two topics.

One way of representing a document (in the case of discussion threads, each post is one document) is as a bag-of-words, *i.e.*, a list of words that appear in the document along with counts of how many times each word appears. Thus, a corpus is represented as a document-word matrix, where each row in the matrix is a single document. As the number of posts in a thread increases, this document-word matrix becomes highly multidimensional; a thread with only a couple dozen posts can contain several hundred unique words. LSA transforms this complex document-word matrix into a simpler representation using singular value decomposition (SVD). This approach decomposes the document-word matrix into several smaller matrices. One result of this decomposition is a matrix of singular values, which can be used to derive a smaller approximation of the original document-word matrix by conflating into a single dimension groups of words that often appear together. These weighted groups of words then become the topics identified by the LSA.

In practice, a number of standard preprocessing steps (cf. 30) are applied to the document-word matrix before performing LSA. The following paragraphs detail these steps.

First, we removed stopwords, which are highly common and generally uninformative words. Stopwords included, among other words: articles, such as “the;” prepositions, such as “about,” “from,” and “over;” common verbs, such as “is,” “have,” and “got;” pronouns, such as “he,” “she,” and “it;” and truncated contractions, such as “couldn,” “doesn,” “hasn,” and “wasn.” To this standard list, we added a number of custom stopwords, including “edu,” “email,” “university,” and “faculty,” which came mostly from signatures that were not automatically removed.

Second, all words were lemmatized, *i.e.*, transformed to their uninflected form. For example, the words “swimming,” “swam,” and “swum” were all converted to “swim.” This lemmatizing uses a part-of-speech tagger, which determines whether a given word is likely to be a noun, verb, adjective, and so forth, since different parts of speech might need to be lemmatized differently. For example, “meeting” can

either be a noun that refers to a group of people who have assembled at a scheduled time for a specific purpose, in which case the lemma is "meeting," or it can be a verb that refers to enacting such an assembly, in which case the lemma is "meet."

Third, words that appeared in very few or very many of the documents being analyzed were filtered, as highly common as well as highly uncommon words can both bias calculations and be uninformative from a topical perspective. Based on our subjective judgment, the implementation described here filtered words that appeared in fewer than 2.5% or more than 50% of documents in a corpus (see 31 for a similar example of pruning extremely rare terms). These limits appear satisfactory in our tests, but could clearly be adjusted further based on the demands of analyzing different corpora.

Fourth and finally, counts of individual words were transformed using term frequency-inverse document frequency (TF-IDF). The TF-IDF calculation divided the word counts (*i.e.*, term frequencies) by the number of documents in which the words appeared (*i.e.*, inverse document frequency). Thus, somewhat common words were given lower weightings, and rarer words were given higher ratings. This final TF-IDF weighted lemma-document matrix was then used as the input for LSA.

We explored a variety of options in choosing the corpus on which to run LSA, including: using a single discussion thread as the corpus, with individual posts as documents; using a collection of discussion threads as the corpus, with individual posts as documents; using a collection of discussion threads as the corpus, with each entire thread treated as a single document. Ultimately, we settled on using a collection of discussion threads as the corpus, and treating individual posts as documents, since we wanted both to see how the discussion was changing on a post-by-post basis, and to understand how different threads explored the same space of possible discussion topics. Landauer and colleagues 27 demonstrate using LSA on a very small example corpus where documents are the titles of technical reports, each no longer than ten words. Thus, we were reasonably certain that individual posts would be sufficiently long to be treated as documents.

Furthermore, when applying LSA, we experimented with specifying different numbers of topics, from 4 up to 250 different topics. We found that requesting LSA to generate 10 topics resulted in topics which were individually sensible as being "about" one idea, and which did not contain significant redundancy with multiple LSA topics seeming to be about the same or very related ideas. Other recent topic modeling work has used numbers of topics on a similar order of magnitude, if slightly larger, for example, from 15 to 25 31, 32. In addition, restricting the analysis to 10 topics made analyzing the results practically tractable; reading through descriptions of 250 different topics did not yield significantly more insight than reading through 10 topics. Since the goal is to support human analysis and faculty training, such pragmatic concerns are important.

LSA is not the only available topic modeling technique; other popular techniques include Latent Dirichlet Allocation 33, PLSA 34, and author-topic models 35. However, current implementations of such techniques are rather more computationally intensive than that for LSA. Furthermore, while these alternative techniques may be more technically robust or theoretically sound from a purely mathematical or probabilistic standpoint, the results of this work demonstrate that LSA provides a reasonable topic model that can still be useful and informative without significant computational overhead. We leave exploring the use of other more complex topic modeling techniques in this context as a topic for future research.

G. Projecting the Posts

To highlight relationships among individual posts, we developed a visualization technique to present the results of the analysis in a readily interpretable manner that foregrounds the impact of the facilitator in

shaping the discussion. This visualization (described below) plots the posts in a two-dimensional space based on their scores for each topic. The challenge was transforming the ten-dimensional topic scores for each post into a two-dimensional representation. For this, we used principal components analysis (PCA), a standard dimension-reduction technique (see 18 for another use of PCA in data visualization). PCA is essentially a projection from a higher-dimensional coordinate system into a lower-dimensional coordinate system such that no two dimensions in the resulting project are correlated with one another. For example, any drawing on a two-dimensional piece of paper is a projection of a three-dimensional object. Depending on the perspective used, that two-dimensional projection may show different aspects of that object more or less prominently; a cube viewed from straight on may appear only as a square, but when viewed from an angle reveals more edges and faces, making it clearer that the object is a cube. The projection resulting from PCA ensures that the first principal component (*i.e.*, the first dimension of the projection) accounts for the maximum variance in the data possible by a single dimension, that the second principal component accounts for the maximum possible remaining variance, *etc*. Here, we used PCA to transform the ten-dimensional topic scores for each post into a two-dimensional representation that can easily be plotted.

H. Topic Space Visualization

We used PCA to create a visualization that shows the conceptual distance between different posts. In this visualization, each post is plotted as a point labeled with a number indicating order in the discussion; the initial post is numbered 0, the first reply is numbered 1, *etc*. Instructor posts are denoted by red solid squares, while student posts are denoted by hollow blue diamonds. The initial post is also indicated by a large red X to make it readily visible. Optionally, a line may be drawn starting at the 0th post that continues through each post in order; the examples presented in this report omit this line.

Because the axes resulting from the PCA have no intrinsic meaning, these visualizations omit axis labels. Instead, we generated a topic-based and a term-based key for the topic space visualizations. The topic-based key gives the clearest insight into how the ten dimensions are related when reduced to two dimensions on the graph. Conversely, it shows the topic ambiguity for a post in that region of space: Two posts close to each other in the same region may actually reflect two distinct topics. Note that posts on the opposite side of the plot from a topic may be interepreted as being the inverse of that topic. This key is prescriptive, dictating where posts are plotted on the graph. The term-based key plots the most important terms according to the topic model; here, we use the 50 most important terms, but an arbitrary number can be specified. This key is descriptive, indicating which terms are associated with which portions of the topic space.

For the sake of clarity, numerical labels are omitted from the axes on the individual discussions' topic space visualizations, with each tick mark representing an interval of 0.01. While the analyses focus on within-discussion patterns rather than between-discussion relationships, the uniform axis presentation still enables comparison to the topic keys.

I. Comparing Quantitative with Qualitative Analyses

Finally, in our qualitative analyses of these discussion threads, we focused on examining the content and discourse patterns in the discussions. With regard to content, we documented the topics addressed, paying particular attention to the depth of discussion around biology topics relevant to the initial discussion question. Our interest in discourse patterns centered primarily on the instructor's facilitation behaviors: the use of questions and declarative statements, responses to students' contributions and questions, attempts to elicit more information and further participation from students, and probing for student understanding. We were also interested in noting the extent to which the instructor refined or corrected students' ideas, elaborated on information in the discussion, and extended the ideas raised by applying them to new contexts.

III. RESULTS AND DISCUSSION

Our goal in choosing case studies was to reflect the breadth of variability in instructors' facilitation strategies and to examine how evident these behaviors were in the visualizations and analyses used here. From the 13 discussion threads available from our search, we selected at least one from each instructor to feature as a case study. This was typically the longest discussion, to capture a wider range of variability in posting behavior. Each instructor displayed a consistent facilitation style across all classes, except for some differences in the timing of posts. We included two threads (N4 and N5) by the same instructor since they exhibited slightly different patterns in timing as well as content. Since the other threads more closely resembled the case studies selected, we omit them from the results reported here. This section presents results from six case studies, including two of the seven discussion threads about Lamarckian evolutionary theory (L3 and L5) and four of the six discussion threads about the mechanisms of natural selection (N2, N3, N4, and N5).

A. Lamarck Threads

1. Topic Model

Table 1 depicts the term weights for the top 20 terms for each of the 10 topics in the LSA-derived topic model. This table also includes the weights for topic 0, which is not a topic per se, but rather indicates the most important terms across all topics; topic 0 also tells which terms are most influential in determining the topics that constitute a particular document.

Some of these topics are rather general, such as Topic 1, which seems to discuss the "differences" between "Lamarck's" and "Darwin's" "theories," particularly with respect to "species," "organisms," and "natural" "selection." Other topics are more specific, such as Topic 3, which appears to pertain to certain "genetic" "codes" leading to a particular "trait" among "offspring," but not specifically with regard to "viruses." There is also partial overlap between some topics. For example, both Topics 6 and 8 discuss "viruses," but Topic 6 focuses more on whether "genes" "can" "change," while Topic 8 seems to pertain more to how certain "traits" "help" certain "generations."

As previously noted, the text preprocessing incorporated various methods for removing extraneous or personally identifying information from the corpus being analyzed. In a few cases, individuals' names were not automatically removed and thus appeared in the topic model. These names have been replaced with "<name>" in the results reported below.

This topic analysis forms the basis for the following discussion thread visualizations. As a representation of the distributions of words that naturally emerged in the seven discussion threads analyzed, it provides an approximate sense of what students were discussing. Thus the terms and weights for this topic model serve as a useful reference when interpreting the visualizations.

Table 1. Topic weights and terms from LSA topic model on "Lamarck" discussion threads.

Topic 0		Topic 1		Topic 2		Topic 3		Topic 4	
0.237	lamarck	0.166	propose	0.322	mechanism	0.117	genetic	0.336	genetic
0.224	darwin	0.162	darwin	0.234	virus	0.098	code	0.302	code
0.205	theory	0.161	lamarck	0.212	inheritance	0.086	immune	0.284	immune
0.194	evolution	0.147	question	0.16	question	0.083	offspring	0.26	system
0.169	change	0.146	difference	0.099	herpes	-0.081	turn	0.254	cold
0.162	trait	0.143	week	0.082	part	-0.084	gray	0.144	question
0.161	species	0.129	day	0.081	outside	-0.084	reproduction	0.141	drink
0.146	offspring	0.122	theory	0.078	genetic	-0.091	non	0.119	affect
0.139	pass	0.117	response	0.078	non	-0.093	problem	0.115	change
0.139	propose	0.114	inheritance	0.076	code	-0.094	discussion	0.114	inheritance
0.131	believe	0.106	discussion	-0.075	turn	-0.1	normal	0.101	predisposition
0.131	organism	-0.111	cold	-0.077	hair	-0.117	outside	0.099	actually
0.124	difference	-0.113	think	-0.084	gray	-0.144	gene	-0.097	pass
0.12	virus	-0.122	pass	-0.088	interesting	-0.152	example	-0.11	offspring
0.118	characteristic	-0.132	gene	-0.09	theory	-0.177	virus	-0.12	mean
0.115	mechanism	-0.134	down	-0.092	good	-0.19	reat	-0.126	will
0.111	selection	-0.136	can	-0.119	<name>	-0.324	thanks	-0.139	cure
0.109	natural	-0.169	code	-0.264	example	-0.325	question	-0.165	herpes
0.108	inheritance	-0.183	genetic	-0.351	great	-0.33	nheritance	-0.213	part
0.107	environment	-0.322	virus	-0.496	thanks	-0.426	mechanism	-0.233	virus

Topic 5		Topic 6		Topic 7		Topic 8		Topic 9		Topic 10	
0.263	skin	0.253	skin	0.153	part	0.17	<name>	0.185	good	0	can
0.21	parent	0.199	mechanism	0.147	virus	0.145	anyways	0.182	thanks	0	skin
0.188	good	0.185	thanks	0.143	herpes	0.131	interesting	0.152	us	0	anyways
0.181	wouldn	0.169	color	0.141	thanks	0.12	think	0.136	wouldn	0	dna
0.17	color	0.151	great	0.131	genetic	0.096	pollution	0.127	give	0	turn
0.168	down	0.138	inheritance	0.13	good	0.094	great	0.114	species	0	hair
0.123	characteristic	0.131	yes	0.125	discussion	0.091	somewhere	0.114	start	0	cure
0.121	personal	0.128	personal	0.114	code	-0.091	pigment	-0.107	question	0	gray
0.12	yes	0.12	example	0.113	post	-0.103	personal	-0.11	ye	0	yes
0.116	sing	0.12	eye	0.109	pass	-0.103	gene	-0.115	example	0	color
0.115	nice	-0.116	post	0.108	week	-0.109	color	-0.118	yes	0	personal
0.113	eye	-0.125	prove	-0.109	still	-0.129	part	-0.119	personal	0	pollution
0.112	give	-0.13	gg	-0.111	hair	-0.13	block	-0.124	great	0	part
0.112	post	-0.135	part	-0.112	turn	-0.161	skin	-0.134	cure	0	somewhere
0.11	anyways	-0.139	dna	-0.118	mechanism	-0.176	yes	-0.144	theory	0	think
-0.115	gene	-0.14	discussion	-0.118	somewhere	-0.177	cure	-0.144	part	0	eye
-0.138	great	-0.14	cure	-0.12	gray	-0.237	hair	-0.154	color	0	court
-0.146	example	-0.141	day	-0.129	pollution	-0.24	gray	-0.225	interesting	0	extinction
-0.152	thanks	-0.142	court	-0.136	age	-0.247	turn	-0.231	skin	0	block
-0.237	virus	-0.179	virus	-0.193	anyways	-0.308	can	-0.286	<name>	-1	hilarious

2. Topic Space

The topic space visualizations for these threads use PCA to project the ten-dimensional topic scores for each post onto two dimensions. Here, the first two principal components from the PCA, *i.e.*, the components used in these visualizations, accounted for 21.2% of the variance in the data. Thus, while these visualizations may be informative, they do not show the entire picture.

In order to help comprehend these topic spaces, we present two keys that show what different regions of the topic space mean. First, *Figure 1* shows where the different topics lie in the space, in the topic-based key. Posts that appear to the far right will have strong scores for Topic 1 and, to a lesser degree, Topic 2; posts on the far left, for Topics 6 and 7; posts toward the top, for Topics 8, 9, and 4; and posts towards the bottom center, for Topic 10. This key dictates how posts are placed in the space, but it does not go very far to provide an intuitive description of the space without constant reference back to the topic model. To

that end, *Figure* shows where terms highly relevant to the topic model appear in the topic space, in the term-based key. While this figure provides an intuitive description of what different portions of the space mean, it does not necessitate that posts containing a particular term (*e.g.*, “theory”) will always be placed in the same region on the graph (*e.g.*, off to the right along the horizontal axis).

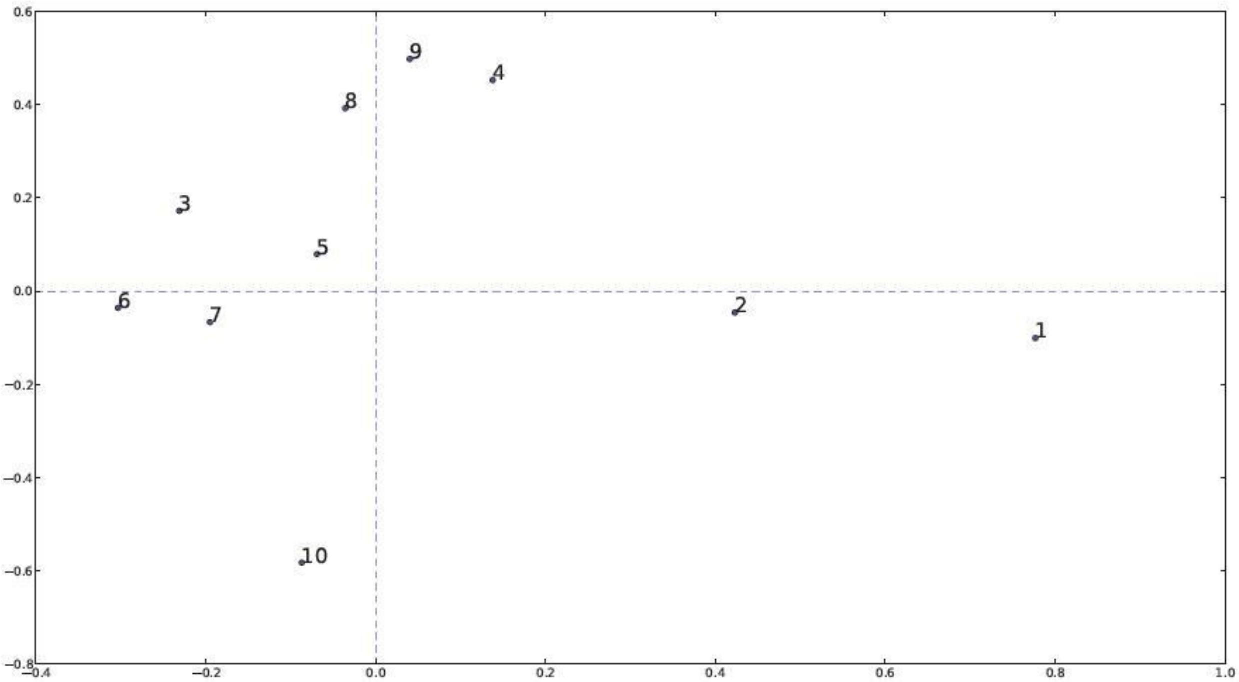

Figure 1. Topic-based key to the Lamarck topic space visualizations.

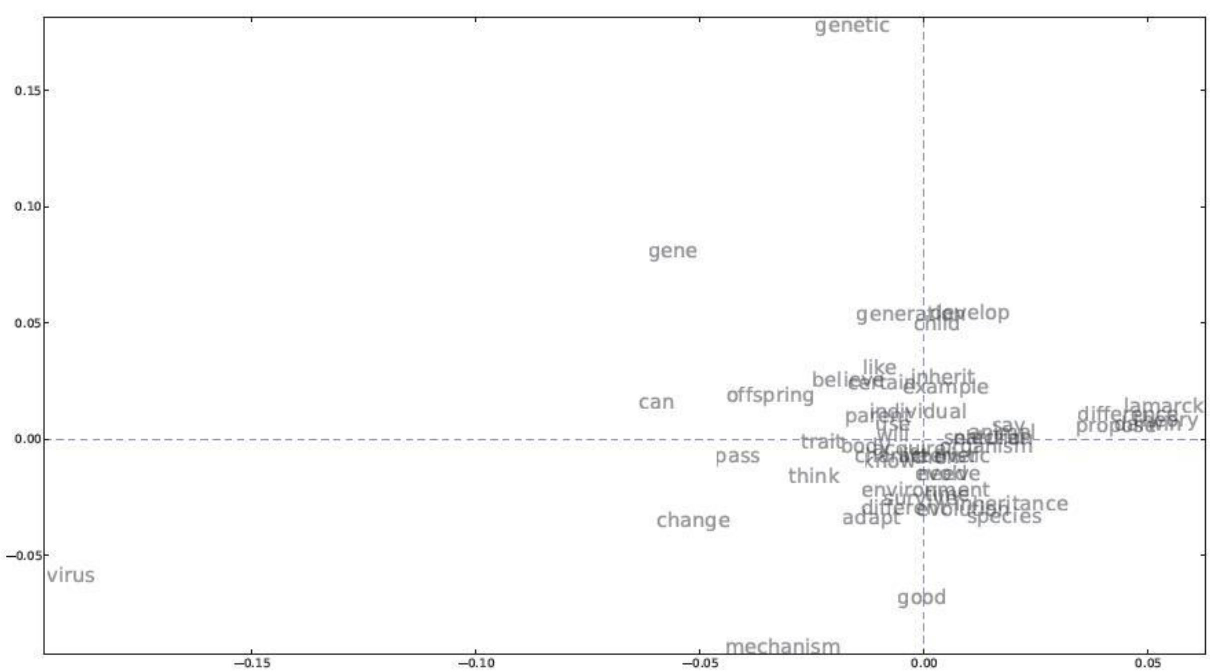

Figure 2. Term-based key to the Lamarck topic space visualizations.

An important point to consider about these topic space plots is that they do not capture all the variability of the data. In the two data sets analyzed, the first two principal components accounted for roughly 25% to 30% of the variability in the topic scores of the posts. There was still a significant amount of data lost in the projection; many posts may actually protrude “into” or come “out of” the two-dimensional plane into which they have been projected. While these visualizations can make it easy to see large patterns within the discussion, one must keep in mind that they cannot tell the whole story.

a. Discussion Thread L3, by Instructor F: Minimal instructor activity

The metrics and visualizations for this discussion thread provide a valuable baseline for what a student-only discussion may look like, since Instructor F did not intervene at all after posting the opening discussion question:

> *Both Lamarck and Darwin understood the importance of inheritance to species evolution. However, there is a subtle but critical difference between the theory proposed by Jean Baptiste Lamarck in the early 1800’s and that proposed by Charles Darwin some 50 years later. In your own words, describe the differences in their theories.*

Some simple quantitative metrics on posting behaviors reveal that students were moderately active in the discussion, averaging 3.06 posts each and 108 words per post (330 words per student in the entire discussion). However, these metrics do not show the content of the discussion, the quality of students’ thinking, or the nature of the interactions.

The first half of the discussion (~28 posts, almost all within the first two days) consisted mostly of students’ initial responses to the question, which were largely a paraphrasing of facts and explanations drawn from other resources (presumably the textbook; other than that, students did not cite their sources here). The later part of the discussion became more personal and less scientific, with students mostly debating whether talent reflected nature or nurture and making their case through anecdotes. Although the original impetus for this topic did relate to the fundamental Lamarckian belief that acquired traits may be inherited, the discussion tended to focus on whether these abilities are inherited or learned, not whether the learned traits are then inherited (as Lamarck had claimed). These posts were also shorter, averaging ~50 words in the second half as contrasted with ~170 words in the first half.

Another pattern evident from examining the active posting time depicted in *Figure 2* is the inequity in participation. Instead of contributing the required minimum of two posts, the instructor and eight students posted only once each, with the remaining ten students carrying on the rest of the discussion.

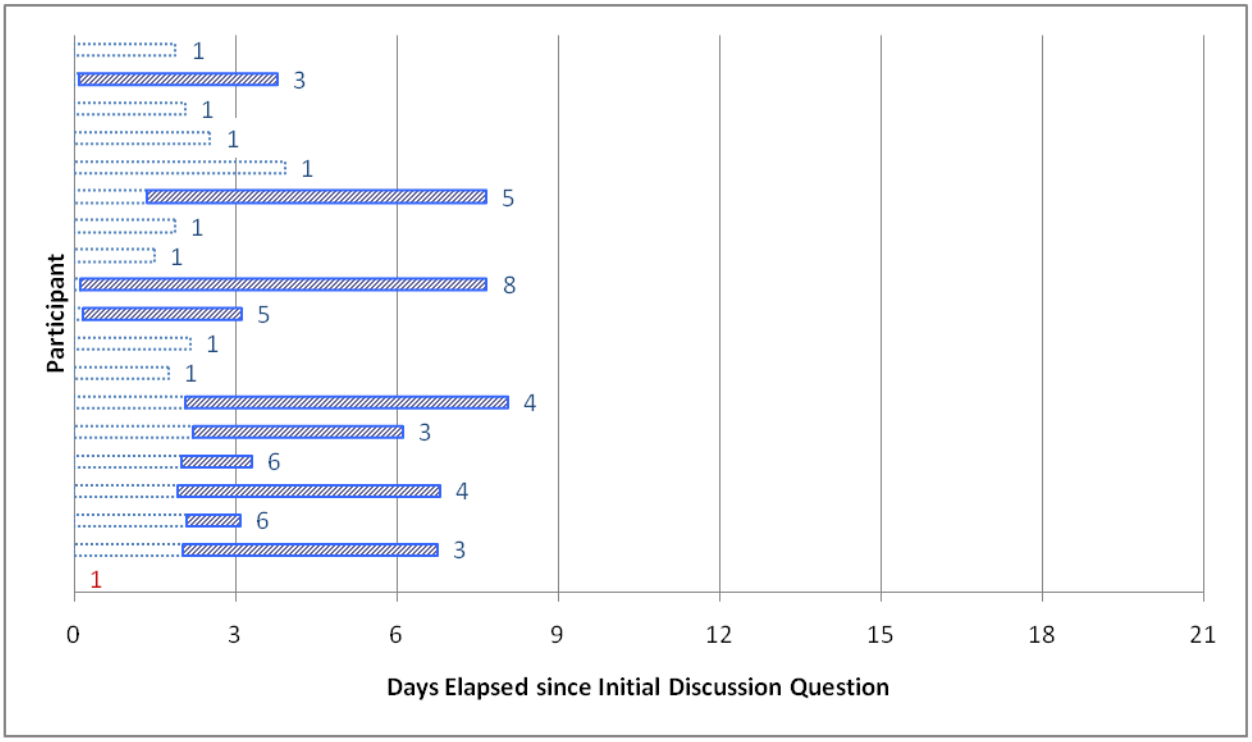

Figure 2. Time interval from earliest to latest posting activity, by participant (blue = student, red = instructor), for discussion thread L3. Solid bars indicate actual posting activity, while blank bars outlined by dots indicate time between initial discussion question and first subsequent post. Labels indicate the total number of posts by each participant.

Students occasionally offered reinforcement to their peers (*e.g.*, "Great example!", "I like the way you break things down"), perhaps taking the place of the kind of encouragement that they might have expected from the instructor. At times the discussion was somewhat dominated by a student who contributed contentful answers but also made jokes and veered off topic occasionally. Some students mentioned that they found it difficult distinguishing between Lamarck's and Darwin's ideas, while others correctly noted that Darwin built upon Lamarck's ideas but did not capture the key difference about the inheritance of acquired traits. These characteristics suggest an unfilled role or missed opportunities for the instructor to intervene and provide more guidance to the discussion. We considered this a low-quality discussion due to the amount of off-topic chatter and the lack of clarity around the fundamental concept of whether acquired traits can be inherited.

One of the attributes most readily apparent in the topic space visualization (*Figure 3*) is how distant the instructor's initial post is from the rest of the discussion. A similar pattern appears for the other three discussion threads from this instructor in this analysis. Superimposing the topic space visualizations for all of the Lamarck threads reveals that it is the initial discussion question that is isolated in the topic space, not the students' discussion. This may reflect the distinctive wording of the question, which also included specific instructions about when and how to post responses, but more importantly, which invited students to generate the difference between Lamarck's and Darwin's theories without including any hints as to what that difference may be. The topic space visualization thus reveals the different language used between this instructor's question and the rest of the discussion.

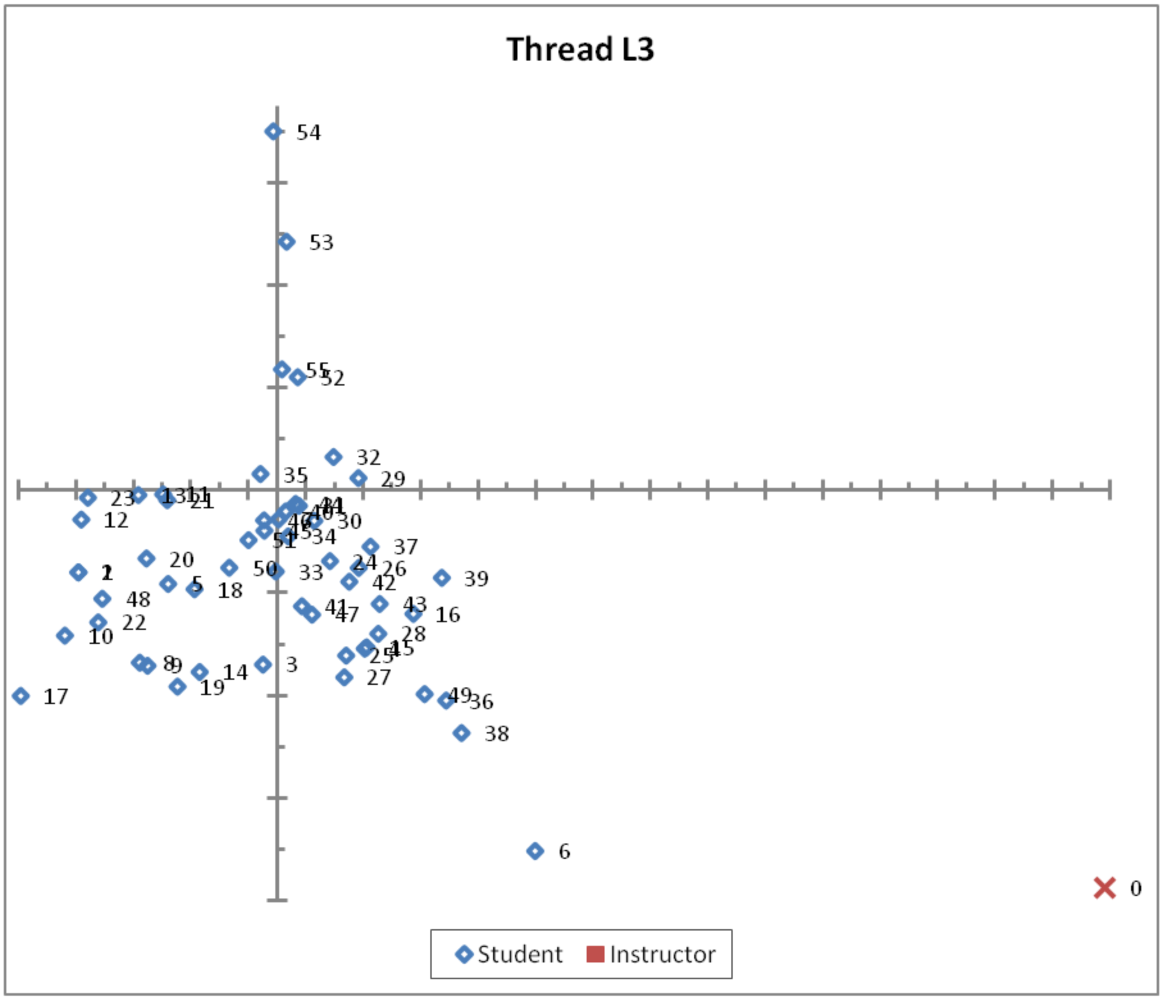

Figure 3. Topic space visualization from discussion thread L3, about the difference between Lamarck's and Darwin's theories of evolution.

b. Discussion Thread L5, by Instructor M: Frequent, continual probing, with high overlap

In contrast, the other instructor whose discussion threads about Lamarckian evolution we included in these analyses was very active in the discussions, yielding different patterns in the visualizations. These discussions addressed the question:

> *Explain what is wrong with Lamarck's notion of evolution. Be aware that I am a bit of a Lamarckian at heart and believe that viruses are a mechanism for evolution!*

Simple quantitative measures of students' posting behavior show very little difference between this thread (L5) and the previous thread (L3). The 17 students in L5 posted slightly more frequently (62 total posts, average 3.65 per student) than the 18 students in L3, with slightly shorter posts on average (94 words *vs.* 108 words). The instructor's posts were comparatively short, averaging 25 words each, accounting for 15.6% of the total word count for the entire discussion. These metrics nevertheless do not capture meaningful differences between the discussion threads in their content or nature of interaction.

Visualizations of temporal patterns in posting behavior can offer some insight into participants' activity (cf. 21). *Figure 4* depicts the word count of each post as a function of the time elapsed since the initial discussion question that opened the thread, with posts by students and by instructor denoted in different colors. This graph reveals an early burst of activity by the students between days 2 and 5, with slight

involvement from the instructor, followed by an extended discussion containing shorter posts by both instructor and students from days 6 through 20.

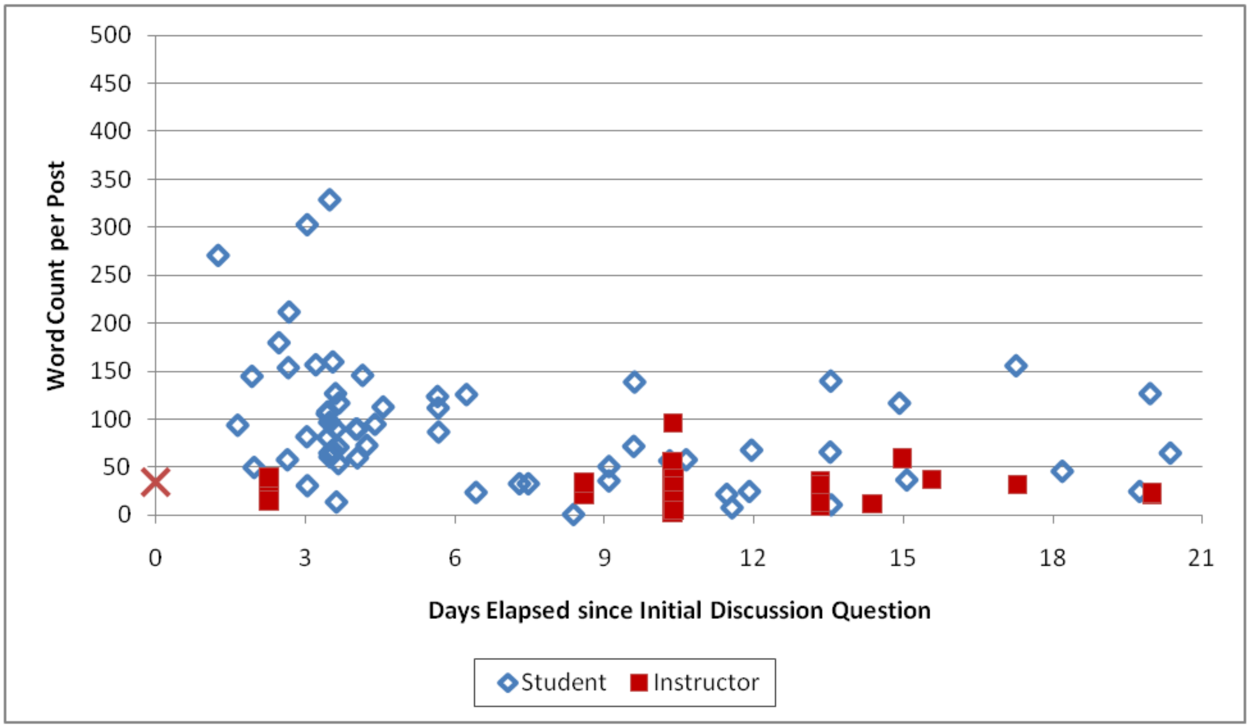

Figure 4. Post length as a function of participant (student vs. instructor) and time (days elapsed since initial question) for discussion thread L5.

Figure 5 reveals tremendous variability among students in their active posting time, with eight students ceasing to post after the initial burst of activity subsided, four continuing approximately through day 12, and three sustaining the discussion along with the instructor through day 20. Since this graph represents posting times rather than forum login times, it provides a conservative view of when participants were active in the discussion.

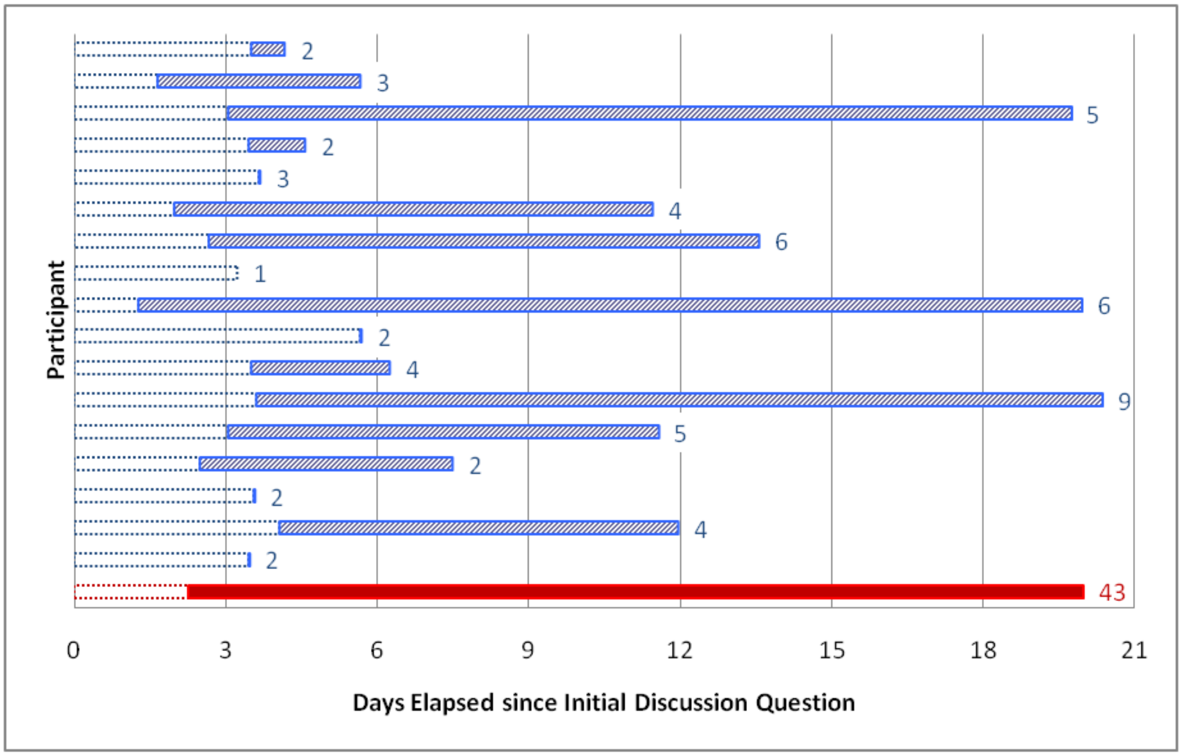

Figure 5. Time interval from earliest to latest posting activity (except for initial discussion question), by participant (blue = student, red = instructor), for discussion thread L5, labeled with each participant's total number of posts.

Most of the discussion in this thread remained on-topic. Those segments that wandered slightly away from Lamarckian theory explored other closely related concepts in genetics and evolution such as chimerism *vs.* conjoined twins, skin color as an adaptation to sun exposure, the evolutionary history of lactose intolerance, and heredity of gluten allergies. Overall, compared to the previous thread, more students demonstrated thoughtful reflection on Lamarck's beliefs, volunteering opinions about the flaws in his theory, pinpointing its problems, and pondering its implications. Students discussed the importance of the role of DNA in heredity, rather than simply talking about the inheritance of physically observable traits. They also identified sources to support and enrich their arguments, with seven students including links to related science, health, and general news sites.

Instructor M's participation after the initial discussion question consisted of multiple (42) short posts across a range of response types. Declarative statements included acknowledging students' contributions, confirming or reiterating the most important idea in a student's post, refining imprecise or partially-correct student assertions, addressing questions or misconceptions, focusing attention on the key concept (mechanisms of inheritance), and introducing new information on the role of viruses in inheritance. Interrogative statements (in 16 posts) included asking students to look up and share answers to specific questions and posing various "what-if" questions. *Table 2* presents some brief examples of each response type.

Table 2. Examples of interventions by Instructor M in Thread L5.

Description	Example
Acknowledge students' contributions	*I think you've crystallized an interesting point here...*
Confirm most important idea	*Yes. The point is that there is no mechanism...*
Refine imprecise or partially-correct claim	*Are these heritable changes though, or just products of good healthcare?*
Address questions / misconceptions	*Probably she stopped making lactase and wasn't so much allergic as lactose intolerant.*
Focus attention on key concept	*It really is a question of mechanisms of inheritance.*
Introduce new information	*Viruses actually insert their genetic code into our own... hence changing our genes!*
Solicit contributions from students	*Care to look for a link on 'fossil viruses' and post it for us?*
Pose questions to students	*What if viruses are a mechanism for us to exchange genes without direct contact or offspring?*

While we did not have sufficiently precise or comprehensive assessment data to compare students' learning across classes, some of the qualitative comments in this discussion strongly suggest that students' understanding changed after instructor intervention. In direct response to the instructor's question or comment, four students expressed initial surprise or confusion that indicated they did not know this information previously, followed by an explanation of the phenomenon in their own words. Three students also included a reference to a credible source describing related information. *Table 3.* compiles these examples.

Table 3. Examples of students' learning after instructor intervention in Thread L5.

Student	Initial surprise / confusion	Restatement of phenomenon	Related references
1	*What has me stumped is relating this to viruses as a mechanism for evolution.*	*[Viruses take over living cells to reproduce, and in so doing], they change the DNA of the host.*	textbook
2	*Wow. I didn't see that coming.*	*The environment can affect our DNA and we could pass a revised set of genes to our offspring?*	news reports on science article documenting ancient mechanism of antiviral defense
		In a later post addressed to a classmate: *[We inherit DNA, and learned characteristics don't change our DNA.] However, certain viruses can actually change our DNA at the molecular level.*	
3	*That is so amazing!*	*It is hard to believe that viruses can actually insert their own genetic code changing our genes permanently. ...I am beginning to share your enthusiasm as I am learning the evolutionary gene transfer process.*	
4	*...this is definitely not what I thought we were talking about. It is quite interesting.*	*So the virus injects its genetic code in ours and directs us to make things for it to replicate itself?*	science blog describing evolutionary consequences of viral resistance

We considered this a high-quality discussion due to the clear focus on the central concept (mechanisms for inheriting acquired traits), thoughtful analysis of how these mechanisms would work, reliance on

supporting evidence, and participants' frequently building upon the content of others' comments and questions.

The topic space visualization for this thread (*Figure 6*) reveals close overlap between the students' posts and the instructor's posts, with a slight suggestion of increased instructor posting toward the upper right and at the bottom of the graph. While the posts at the bottom (59, 76) represent the instructor briefly thanking students for their contributions, those in the upper right (68, 70) correspond to comments pointing out the importance of identifying a "mechanism of inheritance" of traits. This suggests the instructor's attempt to lead the students in a particular direction during the discussion, especially considering that four students had previously mentioned a "mechanism of evolution" but none explicitly connected "mechanism" to "inheritance." The importance of such precision is that the weakness of Lamarck's theory is in failing to address how acquired traits can influence an organism's genetic makeup and thereby be inherited by the offspring.

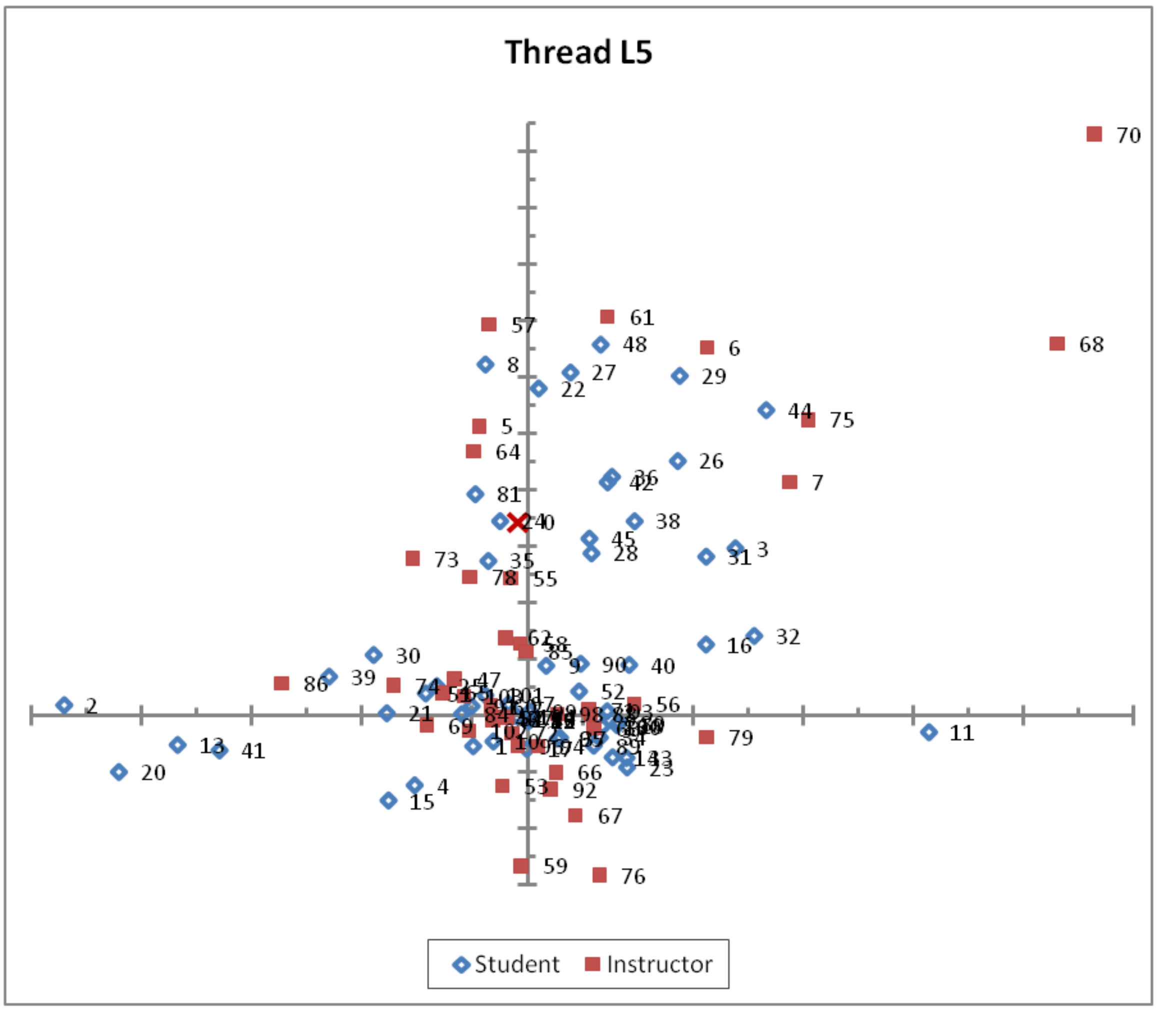

Figure 6. Topic space visualization from discussion thread L5 by Instructor M, about the flaws in Lamarck's theory of evolution.

B. Natural Selection Threads

As the analyses of the foregoing discussion threads indicate, active instructors who participate in the discussions have the opportunity to shape the direction and flow of those discussions. However, simply being active is not sufficient; the discussion threads about natural selection show key differences in the content and nature of interaction that are associated with different outcomes in discussion quality. Although the instructors facilitating these discussions did not post as frequently as Instructor M, their total word count was comparable or greater in most cases. All four instructors also displayed distinctively different facilitation styles, as we will describe.

1. Topic Model

Table 4. shows the details of the topic model for these discussion threads. One might expect to see a topic about Darwin, evolution, and natural selection, but the terms "selection," "natural," "evolution," "theory," "darwin," and others are actually negatively associated with Topic 1. Indeed, no single topic strongly captures these terms. On the one hand, this seems odd, as the ostensible focus of the discussion is natural selection. On the other, it may indicate that, rather than discussing natural selection in a general manner, many posts considered specific instances or applications of the concept. For example, Topic 2 appears to deal with how some "strains" of the "flu" "virus" can "evolve" to be "resistant," both to "flu" "shots" administered by "doctors" and to our "immune" "systems." Somewhat unexpectedly, Topics 3 and, to a lesser degree, 4 seem to deal with "believing" in "Darwin's" "theory" of "evolution" as a "science" and "fact" versus "faith" in "god." Again, a few names escaped detection in the text-cleaning algorithms and have been omitted from the results shown in *Table 4.*. Although it is difficult to tell from the topic analysis alone how closely these topics pertain to the intended focus of the thread, it indicates that they are prominent topics within the discussion. Reading individual discussion threads reveals how these topics play out in the context of the interactions between the students and the instructor.

Table 4. Topic weights and terms from LSA topic model on "natural selection" discussion threads.

Topic 0		Topic 1		Topic 2		Topic 3		Topic 4	
-0.096	survival	0.173	flu	0.74	flu	0.306	theory	0.278	science
-0.103	good	0.162	think	0.429	shot	0.266	science	0.24	fact
-0.104	can	0.155	people	0.202	virus	0.2	fact	0.207	believe
-0.105	role	0.129	like	0.135	strain	0.199	darwin	0.196	faith
-0.108	population	0.116	go	0.134	immune	0.191	<name>	0.153	evolve
-0.111	think	0.107	child	0.117	doctor	0.185	evolution	0.146	god
-0.115	organism	0.103	know	0.084	system	0.17	believe	0.127	will
-0.118	process	0.101	shot	0.079	resistant	0.152	faith	-0.11	online
-0.119	example	0.094	work	0.073	sick	0.143	write	-0.11	<name>
-0.12	survive	-0.102	trait	0.069	bad	0.138	god	-0.114	man
-0.12	darwin	-0.11	organism	0.065	selection	0.137	<name>	-0.115	natural
-0.124	will	-0.115	environment	0.056	type	0.111	<name>	-0.124	share
-0.129	change	-0.124	darwin	-0.055	look	0.108	online	-0.129	information
-0.131	trait	-0.127	theory	-0.055	faith	0.102	<name>	-0.13	selection
-0.136	environment	-0.127	role	-0.058	dog	0.102	message	-0.137	child
-0.138	theory	-0.14	process	-0.058	god	-0.105	color	-0.137	dog
-0.163	species	-0.141	species	-0.07	fact	-0.105	example	-0.146	<name>
-0.19	evolution	-0.158	evolution	-0.072	animal	-0.128	trait	-0.19	job
-0.255	natural	-0.268	natural	-0.093	believe	-0.139	strong	-0.205	<name>
-0.272	selection	-0.307	selection	-0.101	science	-0.169	survive	-0.237	woman

Topic 5		Topic 6		Topic 7		Topic 8		Topic 9		Topic 10	
0.502	woman	0.193	woman	0.388	information	0.219	darwin	0.311	strong	0.172	virus
0.355	man	0.171	information	0.364	share	0.134	family	0.193	woman	0.167	resistant
0.142	show	0.158	job	0.214	thanks	0.132	population	0.165	science	0.142	evolve
0.14	study	0.142	man	0.161	trait	0.126	individual	0.156	fact	0.142	population
0.123	<name>	0.142	share	0.158	job	0.115	trait	0.147	survive	0.114	taste
0.099	humans	0.139	example	0.135	darwin	0.102	write	0.125	business	0.114	insect
0.093	longer	0.136	good	0.132	dog	-0.1	science	0.116	faith	0.107	trait
0.091	physical	0.084	physical	0.129	class	-0.101	role	0.112	man	0.102	eat
0.084	immune	0.077	show	0.117	good	-0.103	evolve	0.108	share	0.098	generation
0.084	system	0.077	mechanism	0.114	tree	-0.126	dale	-0.102	weight	-0.096	tree
-0.1	class	-0.075	end	0.112	article	-0.137	mechanism	-0.103	type	-0.104	foot
-0.107	natural	-0.076	business	0.112	color	-0.143	virus	-0.106	resistant	-0.106	explain
-0.108	selection	-0.078	individual	0.107	woman	-0.146	breed	-0.131	virus	-0.121	humans
-0.119	work	-0.082	hunt	0.105	individual	-0.147	resistant	-0.139	live	-0.136	larger
-0.122	example	-0.082	theory	0.102	offspring	-0.154	natural	-0.167	use	-0.145	live
-0.13	thanks	-0.093	animal	-0.106	people	-0.158	selection	-0.168	body	-0.151	high
-0.166	good	-0.131	darwin	-0.118	child	-0.173	man	-0.186	high	-0.152	survive
-0.233	share	-0.159	pay	-0.136	strong	-0.194	example	-0.189	foot	-0.198	environment
-0.251	information	-0.304	breed	-0.151	natural	-0.22	woman	-0.196	evolve	-0.204	well
-0.254	job	-0.655	dog	-0.162	selection	-0.341	dog	-0.226	brain	-0.516	dale

2. Topic Space

Here we present example topic space visualizations based on the above topic model. In this case, PCA accounted for 21.2% of the variance in the data. *Figure 7* and *Figure 8* show the topic key and term key, respectively, for these topic space visualizations. These should be read and interpreted in the same way as the topic space keys from the Lamarck discussion threads.

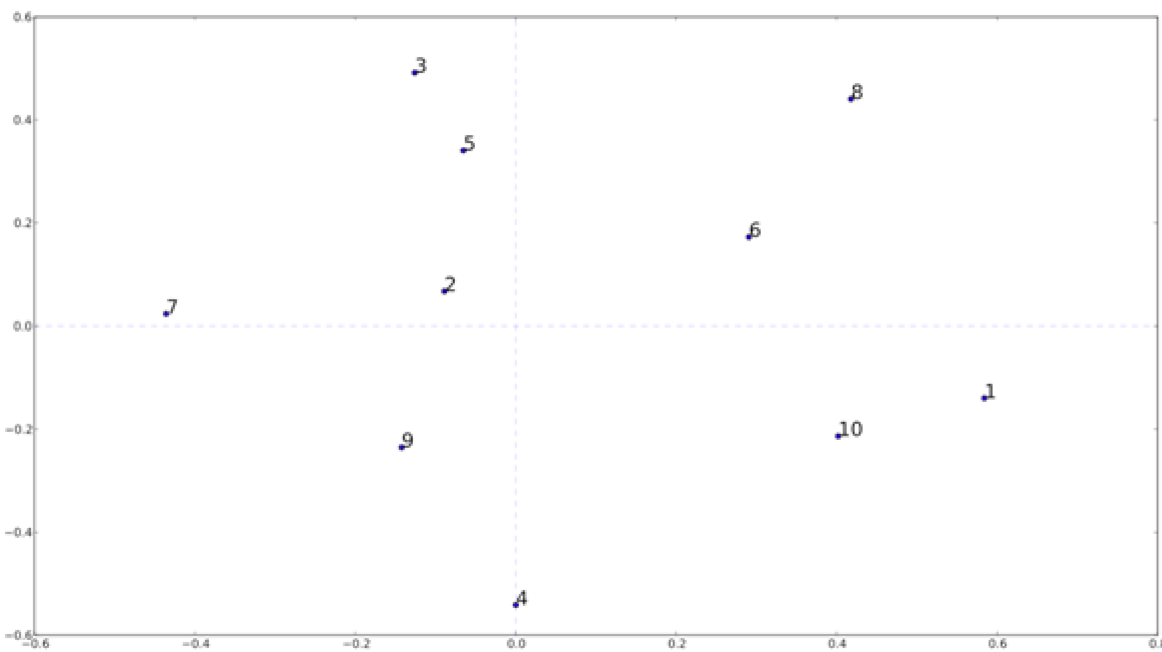

Figure 7. Topic-based key to the natural selection topic space visualizations.

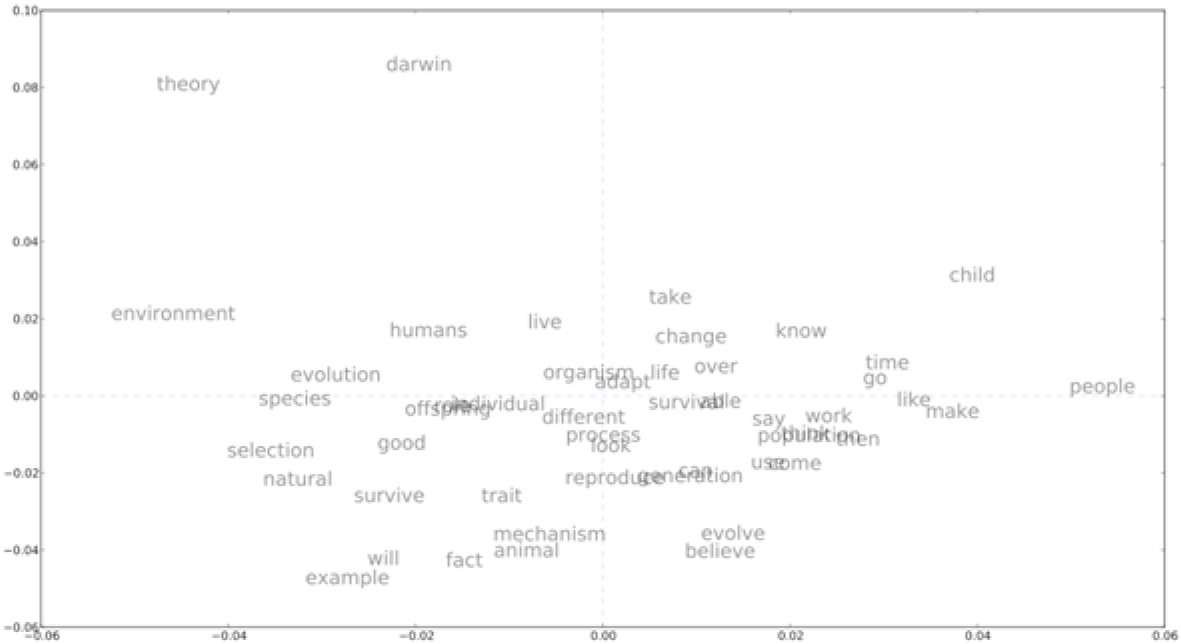

Figure 8. Term-based key to the natural selection topic space visualizations.

a. Discussion Thread N2, by Instructor P: Early, brief instructor requests

This thread illustrates a pattern of activity emerging from a discussion in which the instructor was moderately active at the beginning, using a singularly consistent facilitation style. This discussion addressed the question:

Please describe extensively the role of natural selection in the theory of evolution.

In this discussion, the instructor posted 22 times, while the 17 students posted 116 times, for an average of 6.82 posts per student. The average word count per post was 35 for the instructor and 93 for the students, with the instructor contributing 764 words (6.7%) and the students contributing 10714 words (93%) out of the entire discussion. Thus, this instructor was active but provided a noticeably smaller proportion of the content of the discussion, compared to the other instructors (with the obvious exception of Instructor F, the minimally-active instructor in Thread L3, described previously).

A temporal analysis of the discussion (*Figure 9*) reveals that all of the instructor's participation occurred within the first four days of the discussion. In contrast, almost all of the students (at least 15 of 17) continued the discussion after the instructor's last post, four of them beyond 12 days after the discussion started (*Figure 10*). Coupled with the observations from reading the discussion thread, these visualizations highlight how this instructor influenced the early discussion through frequent and relatively short posts, but did not take advantage of opportunities to intervene in the later discussion in which many students continued to participate.

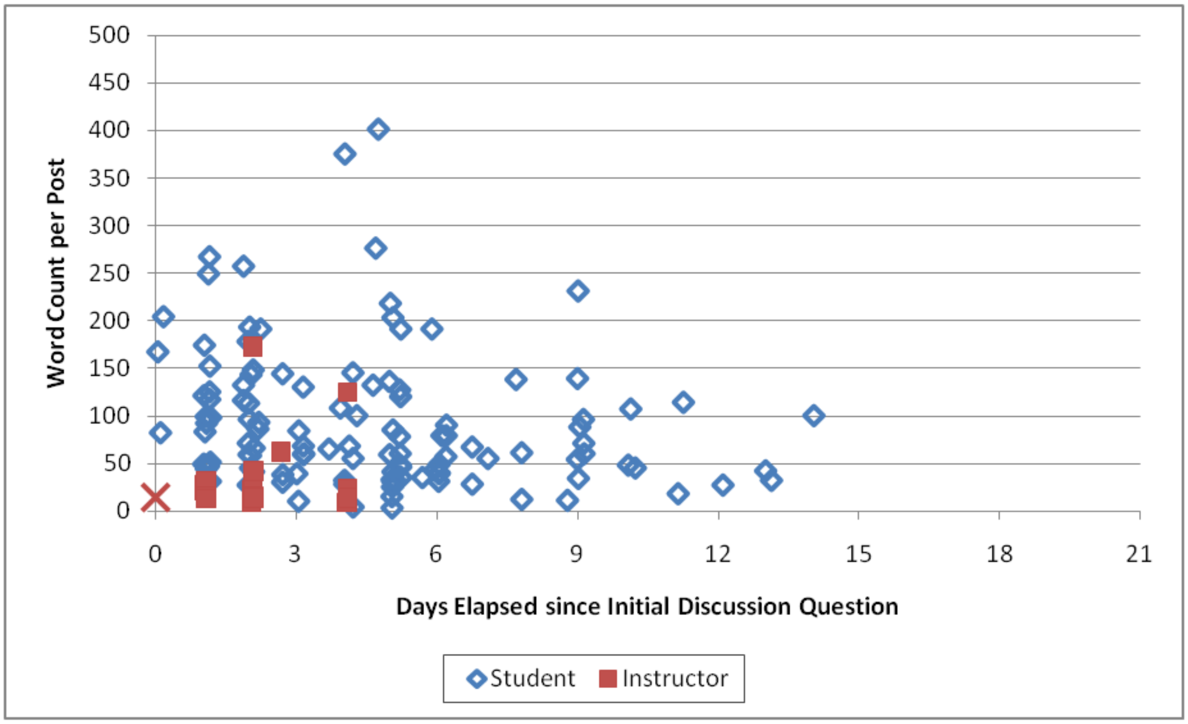

Figure 9. Post length as a function of participant (student vs. instructor) and time (days elapsed since initial question) for discussion thread N2.

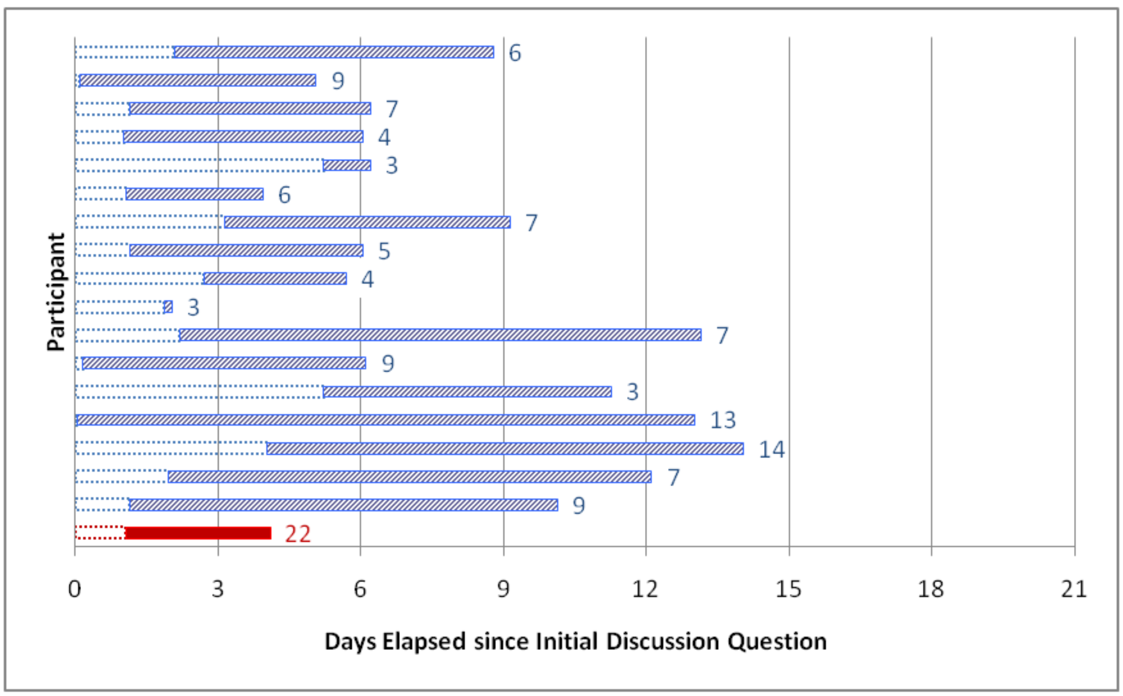

Figure 10. Active posting time (in days elapsed since initial question) as a function of participant (blue = student, red = instructor) for discussion thread N2, labeled with each participant's total number of posts.

As in discussion thread L3, the students here tended to generate textbook-derived answers to the opening question, sometimes examining associated concepts but spending more time talking about tangential anecdotes. Their answers focused a lot on overproduction and competition, less on reproduction and inheritance, suggesting that they may have been parroting explanations in their textbook but not necessarily fully understanding or appreciating the complete mechanism of natural selection. While their

early elaborations generally were relevant to natural selection and evolution, they often sparked longer discussions with personal stories that veered away from the key concepts (*e.g.*, driving at high altitudes, raising twins, animal breeding, effectiveness of flu vaccination). There was also some confusion between inherited and learned traits which did not get directly addressed.

The instructor's participation in this discussion typically took the form of short posts affirming the students' contribution (*e.g.*, "Good job!"), with the occasional request of, "Can you say more?" The positive acknowledgments sometimes identified what the student had mentioned that was a good example, but did not explain why it was a good example or challenge the students to think more deeply about all of the phenomena involved. In one case, the student described how predation creates pressures for organisms to overproduce, without explaining the necessary step of how particular genes and the traits expressed may favor certain organisms to survive, reproduce, and thereby pass along those favored genes. In another case, an accurate statement about how humans and chimpanzees split apart on the evolutionary tree became misinterpreted as a claim that humans evolved from chimpanzees, which one student further misunderstood and subsequently used as a basis for questioning evolution. This then turned into a discussion of creationism as a legitimate alternative to evolution. The instructor did not intervene in this portion of the discussion.

Overall, it appeared that the instructor was primarily focused on getting students to answer the original question and to provide an example of how natural selection works. Three posts provided additional elaboration on the concepts (*e.g.*, mutations in the flu virus, population control in other countries), but were somewhat narrow in focus and did not extend upon natural selection significantly. Otherwise, the instructor did not offer much encouragement for the students to build on their own or each other's ideas. We considered this a medium-to-low-quality discussion due to the prevalence of off-topic personal anecdotes, persistent confusion over fundamental evolutionary concepts, and lack of depth in exploring the process of natural selection.

A quick glance at the topic space visualization, shown in *Figure 11*, suggests three distinct regions: one where only the instructor posts, one where only the students post, and one where both instructor and students post. The instructor-only region consists of the opening question, plus two posts thanking the students for their contributions and asking for an example. The regions where instructor and student posts overlap tend to reflect areas where the instructor's posts included more specific detail about the student post to which s/he was responding. This could take the form of reinforcing what the student said correctly, or elaborating further. The student-only region includes some slightly off-topic discussion as well as some relevant discussion; much of the later discussion in which the instructor did not participate is contained here.

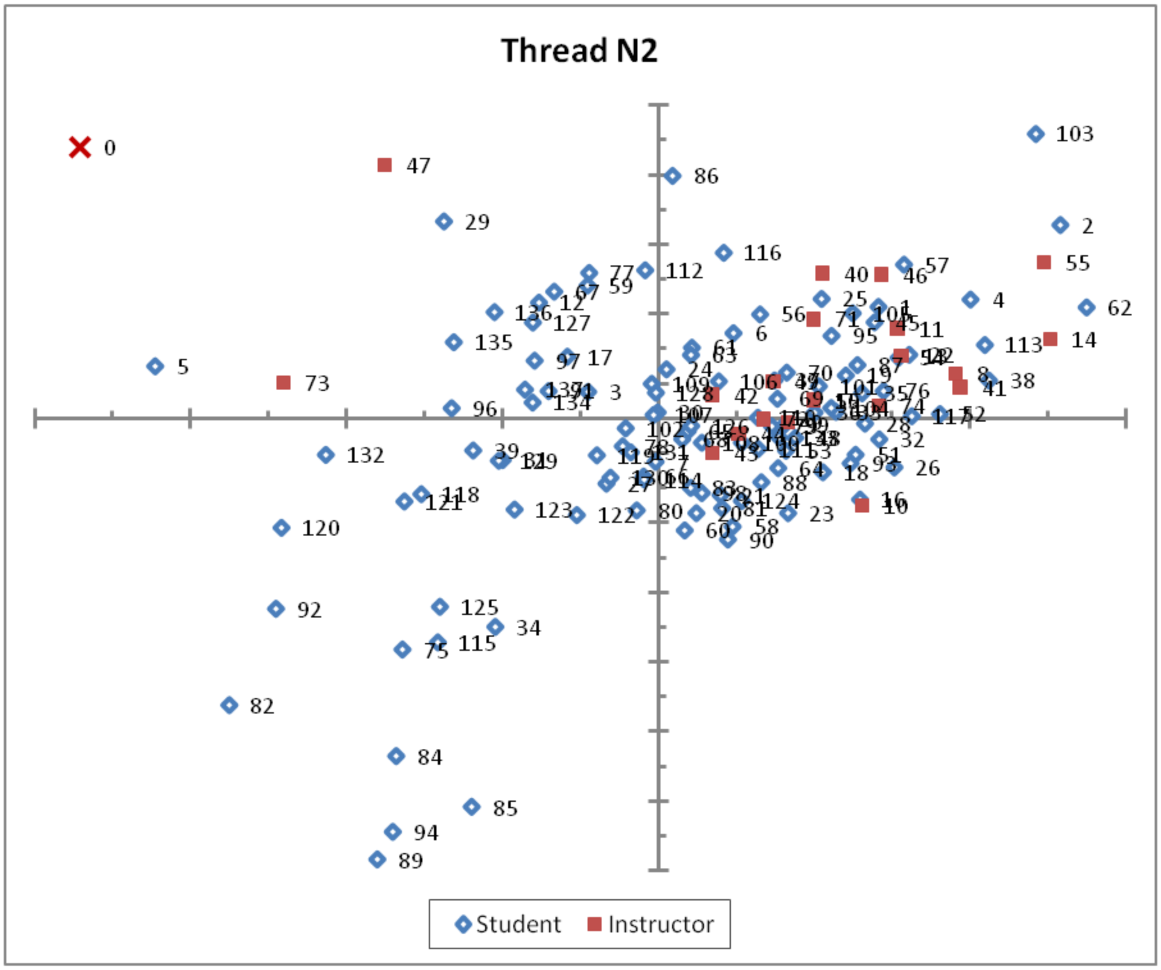

Figure 11. Topic space visualization from discussion thread N2 by Instructor P, about the role of natural selection in evolution.

While the temporal visualizations are helpful in capturing the uneven participation in the discussion, they provide only a partial picture of the discussion's content and interaction. The topic-space visualization reveals more information about the relationship between the instructor's and students' posts, corroborated and enhanced by reading the discussion. Lack of overlap in the topic space corresponds to either complete lack of reinforcement or nonspecific acknowledgment by the instructor, using language that did not resemble the student's post. This suggests that overlap matters, at the very least as an indicator of when the instructor acknowledged students' ideas and encouraged them to explore further. However, while overlap does show where the instructor used similar language to the students, it does not necessarily reveal deep probing. In spite of the apparent topic space overlap for this thread, most of the instructor's posts in that space simply thanked students for what they said and asked for an example, with very little questioning or elaborating further on natural selection.

b. Discussion Threads N4 & N5, by Instructor I: Late, lengthy, off-topic elaboration

In these threads, the instructor asked students to extend the concepts to new contexts and elaborated at length on their contributions, intervening late in N4 in particular. As before, the topic-space overlap shows instructor reinforcement of student contributions, but these areas of overlap may not correspond to the most important topics of discussion. The central question in these discussions was:

> *What is the role of natural selection in the mechanisms of evolution? Provide an example of how this process works.*

Quantitative metrics show that both instructor and students were fairly active in thread N4, with the instructor providing 14 posts and the 15 students providing 89 posts, for an average of 5.93 posts per student. The instructor's posts averaged 138 words, while the students' posts averaged 169 words each, so that the instructor contributed 1925 (11%) and the students contributed 15017 (89%) of the total word count. However, not all of this activity necessarily constituted deep exploration of the key concepts or even stayed on topic. Thread N5 includes more posts but otherwise shows a similar pattern.

In contrast to the previous thread (N2), the instructor did not intervene until six days after posting the initial discussion question in N4, as shown in *Figure 12* and *Figure 13*. During this time, the students were very active in the discussion, although five students did not post again after this interval and may have missed out on the instructor's subsequent intervention.

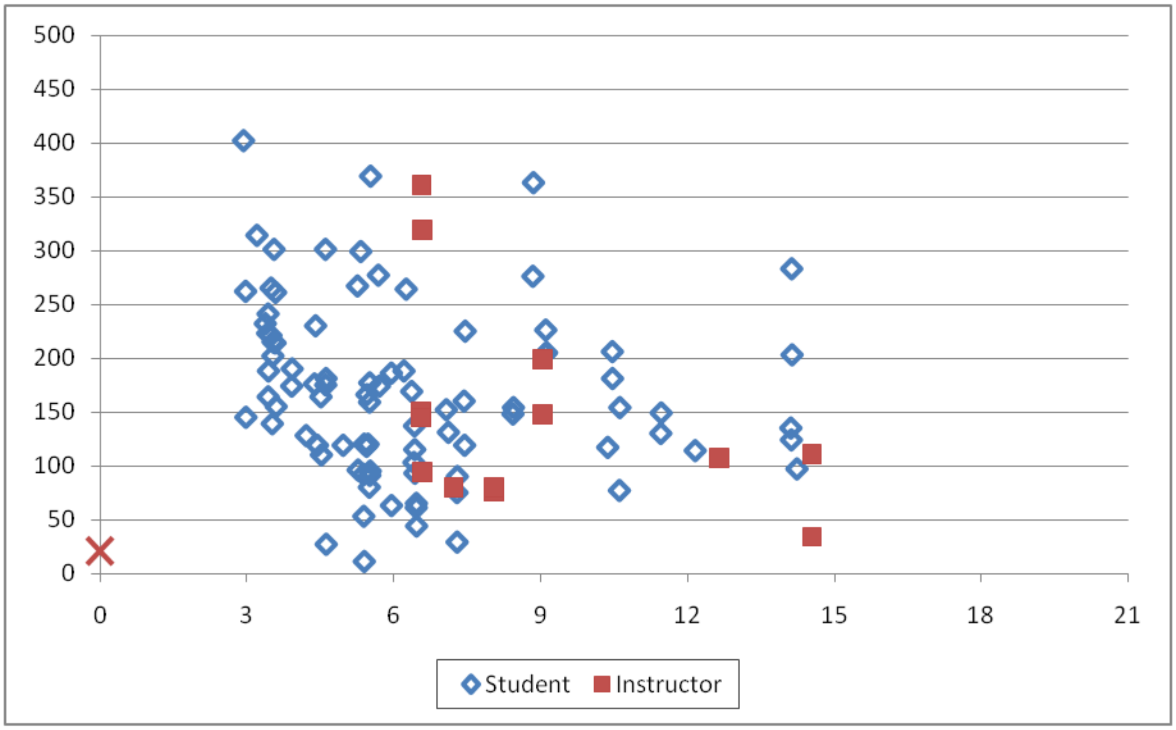

Figure 12. Post length as a function of participant (blue = student, red = instructor) and time (days elapsed since initial question) for discussion thread N4.

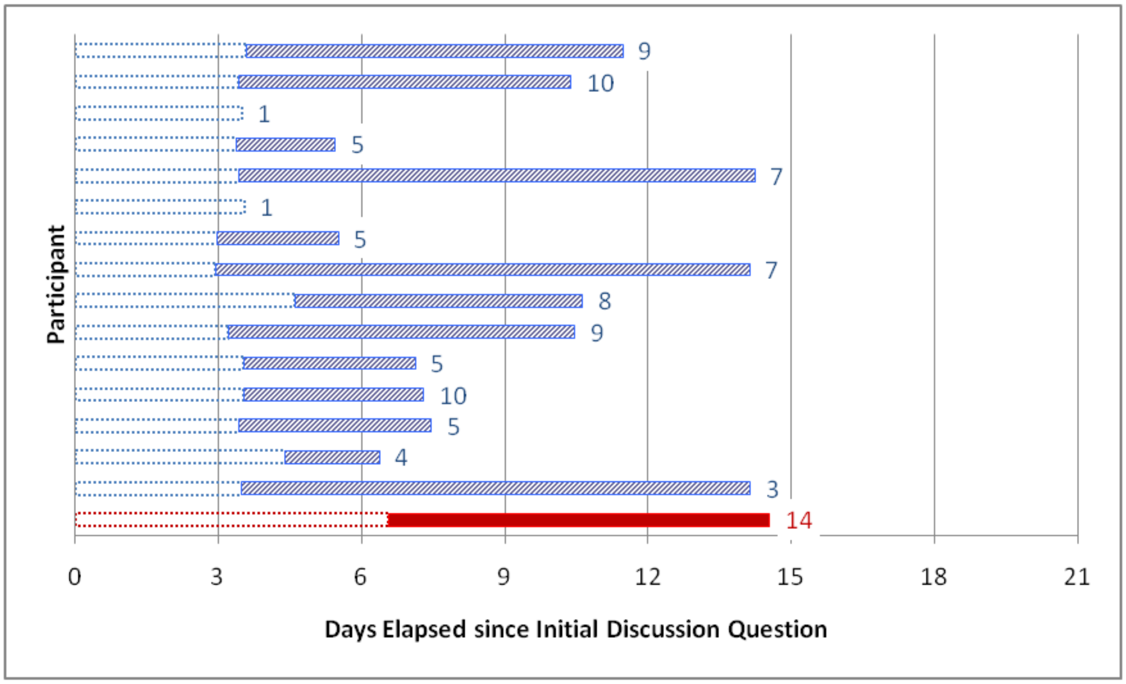

Figure 13. Active posting time (in days elapsed since initial question) as a function of participant (blue = student, red = instructor) for discussion thread N4.

Similar to the patterns in threads L3 and N2, students' initial posts in thread N4 tended to be long paraphrases of encyclopedia-type definitions and recitations of science history, but not so much an analysis of the mechanisms of natural selection. Inspired by the first respondent's comment about people's technology dependence interfering with evolutionary processes, several students began discussing concerns about technology fostering laziness and unhealthy habits. While the original impetus for this discussion explored the implications of altering the process of natural selection through technology, the subsequent discussion focused instead on everyday topics of diet and exercise rather than the evolutionary process. Another subtopic emerged around the efficacy of the flu vaccine, following one student's comment about the ability of the influenza virus to adapt quickly. Again, while the initial comment addressed concepts relevant to evolution, the ensuing discussion focused more on the value of vaccination and personal anecdotes about getting shots and/or getting sick.

As the discussion continued to unfold, another topic emerged in which students were debating the legitimacy of evolution *vs.* creationism. One student expressed doubts about the validity of natural selection based on a combination of misconceptions and flawed reasoning, suggesting that evolutionary theory was a dangerous concept that did not make sense because of the continued existence of monkeys, disagreement about overpopulation and global warming, and confusion over exponential population growth. The instructor very respectfully and carefully addressed these misconceptions by providing multiple sources of evidence demonstrating or explaining the phenomena in question. Still, the continuation of student posts expressing a misunderstanding of what constitutes legitimate evidence and scientific disagreement regarding evolution and creationism, even after the instructor's intervention, suggests both how deep-seated these misconceptions are and how critical it is for instructors to prevent or address them effectively. Having students publicly and repeatedly express resistance to accepting evolution despite instructor intervention may do significant damage to the beliefs of the rest of the class, by appearing to legitimize non-scientific beliefs as holding equal footing to genuine scientific reasoning. We considered this a medium-quality discussion, with the occasional off-topic comments and resistant misconceptions as drawbacks, but the conversation about relevant evolutionary processes and direct addressing of those misconceptions as benefits.

The topic space visualization for thread N4 (*Figure 14*) reveals a fair amount of instructor-student post overlap in the region of heaviest discussion. However, several of these posts represent off-topic comments about dependence on technology, cord blood banking, and 3-D ultrasounds, rather than the more evolution-relevant discussion of antibiotic resistance, genetic drift, camouflage, and exaptation (co-opting previously inherited traits for new purposes). Although the latter concepts all emerged in the thread at some point, they did not experience sustained or deep discussion. In contrast, the instructor's posts that directly address student misconceptions about chimpanzees and global warming (64, 66) appear in a sparser region of the graph.

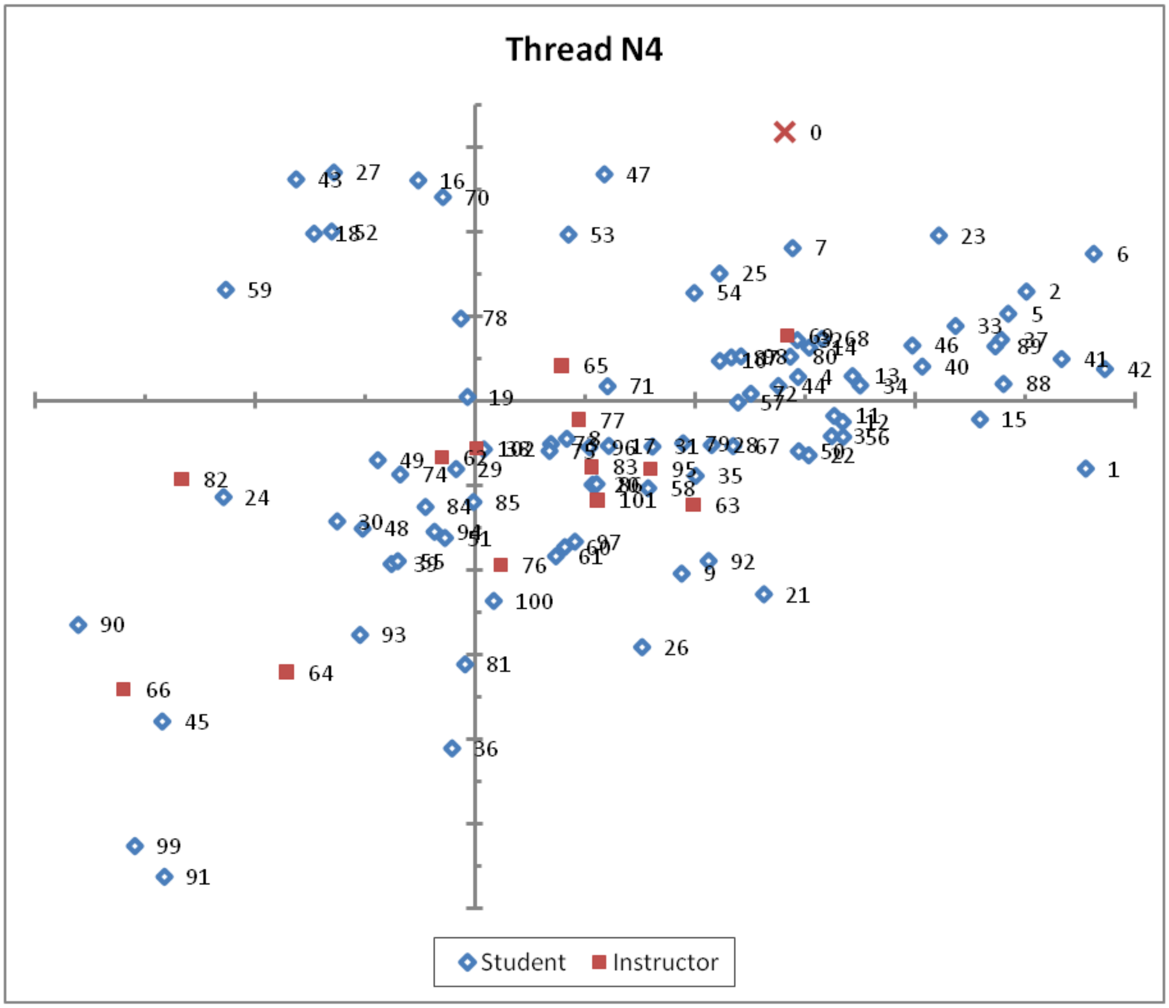

Figure 14. Topic space visualization from discussion thread N4 by Instructor I, about the role of natural selection in evolution.

Discussion thread N5 (Figure 15) showed earlier instructor intervention but a similar pattern in overlap: the area of densest overlap corresponded to off-topic chatter, while the sparsest region of the graph contained the opening discussion question and the students' initial responses (*Figure 16*). In response to the instructor inviting students to apply the concepts of natural selection "to an example outside of nature," much of the ensuing discussion then explored how "survival of the fittest" plays out in the housing market and in employment as a law enforcement officer. These examples are incomplete analogies since they do not include one of the most fundamental yet difficult-to-understand concepts in natural selection: namely, the role of genes and inheritance. On the one hand, relating natural selection to everyday experience may have helped make the general concept more familiar; on the other hand, this may have displaced or prevented a deeper discussion about the relationship between genetics and evolution, and may even have reinforced an imprecise, overly general understanding of the mechanisms

by which evolution happens. The separation between the discussions of more-relevant and less-relevant issues suggests a possible disconnect between students' understanding of concepts fundamental to natural selection and of the more everyday topics discussed elsewhere. We evaluated this as a medium-quality discussion, since it addressed moderately relevant but not the most fundamental concepts in natural selection.

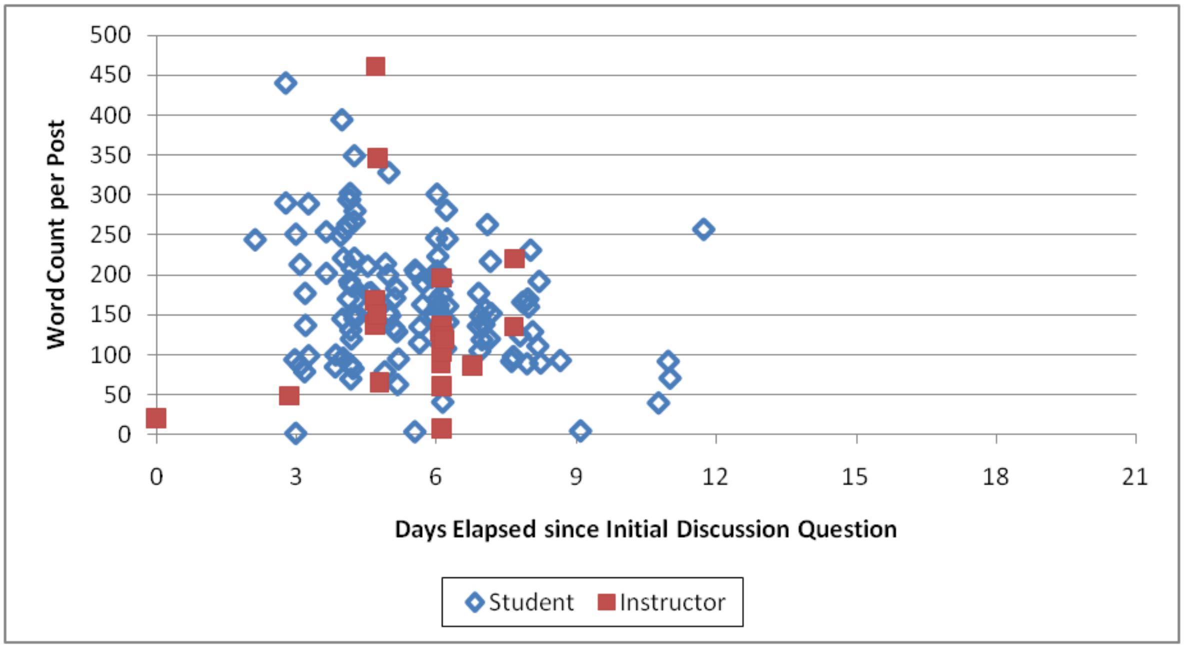

Figure 15. Post length as a function of participant (blue = student, red = instructor) and time (days elapsed since initial question) for discussion thread N5.

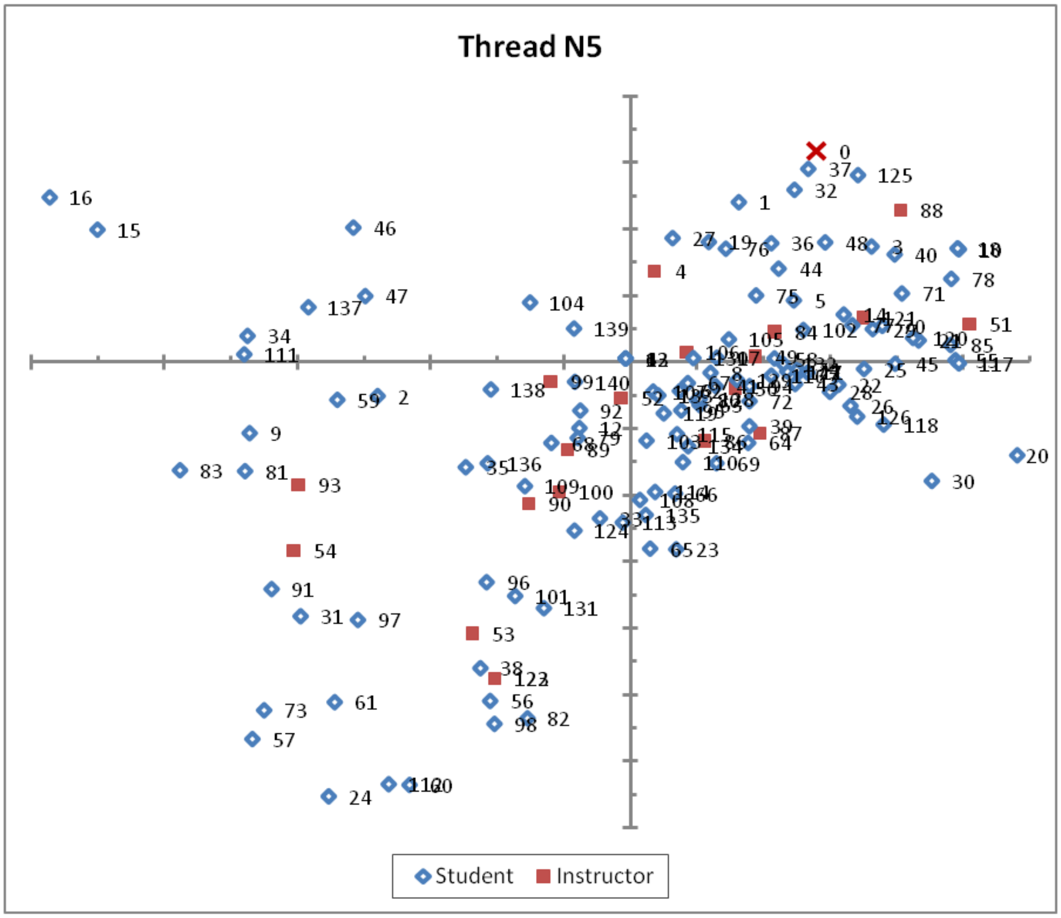

Figure 16. Topic space visualization from discussion thread N5 by Instructor I, about the role of natural selection in evolution.

These analyses indicate that posting length and frequency are not guarantees of discussion quality, nor is topic-space overlap between instructor and student posts. Still, the topic space visualizations suggest that instructor reinforcement does have an impact on influencing what students will discuss. Ultimately, these visualizations do not speak for themselves, but rather facilitate understanding the discussion (*e.g.*, by faculty trainers).

c. Discussion Thread N3, by Instructor B: Sustained elaboration, with low overlap

This discussion thread reveals yet another pattern of activity, with sustained participation throughout the discussion by the instructor, as captured by some of the visualizations. Although this instructor continually posed thoughtful and sophisticated questions for the students to consider, the students seemed to be more engaged by other topics and did not always follow up on the ideas raised by the instructor, a pattern which becomes evident in the topic space visualization.

In this discussion, the instructor provided 18 of the 73 posts, with the 14 active students providing the remaining 55, for an average of 3.93 posts per student. The average word count per post was similar for instructor (76.3) and students (78.3), although the instructor contributed a greater proportion of the total words in the discussion (1373 out of 5677, or 24.2%), compared to the other discussion threads analyzed.

Tracking post length over time by participant type reveals that Instructor B generally remained active throughout the discussion (*Figure 17*). Student post length showed a gradual decrease over time, although slightly less pronounced than in thread N2.

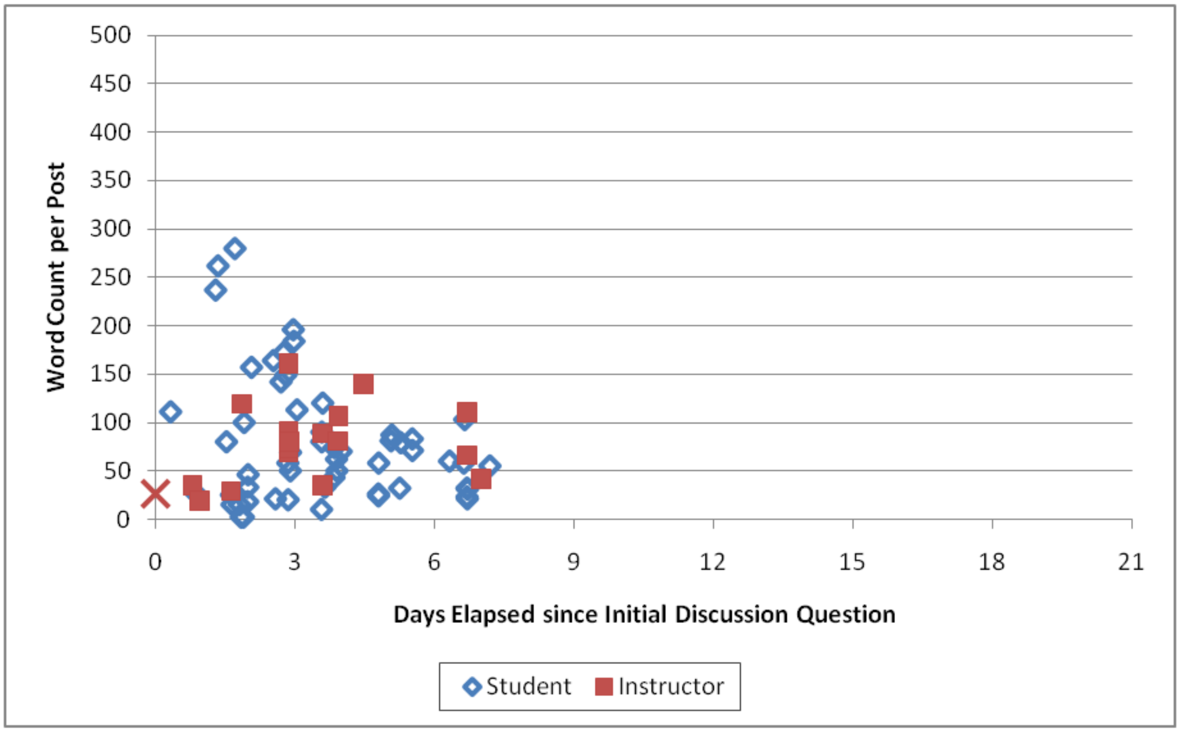

Figure 17. Post length as a function of participant (student vs. instructor) and time (days elapsed since initial question) for discussion thread N3.

In this thread, students began with a short essay-type response answering the initial question, then continued with brief questions to the instructor or follow-up comments elaborating on everyday examples of relevant phenomena that had caught their interest. Although the off-topic discussion was more relevant to evolution and biology than in the other natural selection threads, the students here tended to be more engaged in discussing topics of general interest (*e.g.*, cultural perceptions of beauty) than in exploring the processes of natural selection. The main themes here were: clarifying the involvement of multiple historical figures in the development of evolutionary theory, expressing amazement over discoveries or implications from evolutionary biology, and describing interesting examples of natural selection. While the discussion acknowledged the role of genes in natural selection, it did not explore these mechanisms in detail. One particularly interesting question which the instructor posed was how to explain altruism according to Darwinian theory; although two students did attempt to address this question, ultimately the instructor provided supplementary material explaining this and the subsequent discussion focused on other topics instead.

Instructor B's participation in this discussion generally took the form of introducing new information for the students to consider, such as reminding the students to acknowledge Wallace's historical role in influencing Darwinian evolutionary theory or posing the question about altruism as noted above. In another case, the instructor elaborated on the "slow steps" of evolution that one student mentioned by providing the term "punctuated equilibrium" and explaining it in more detail. Other cases included multiple examples of the influence of sexual selection among humans and other species.

There was a slight disconnect between what the students discussed and what the instructor discussed, with the students talking about more everyday phenomena (except when quoting or paraphrasing other sources) and the instructor explaining more scientific phenomena in technical language. Although some students did provide comments that suggested that they understood the questions, their responses do not provide strong evidence of fully understanding the answers and their implications. We rated this as a medium-to-high-quality discussion due to the exploration of many relevant but not always most

fundamental concepts about natural selection, with some concerns about the disconnect between the instructor's sophisticated commentary and the students' everyday commentary.

This disconnect becomes apparent in the topic space visualization for the discussion (*Figure 18*). The students' posts and the instructor's posts occupy distinct regions of the topic space, with only a couple of student and instructor posts appearing in the "other" region. Thus, the topic modeling emphasizes differences in language used by the discussion participants, with the topic space visualization highlighting the possibility that the students may not have understood the sophistication of the language and ideas expressed by the instructor.

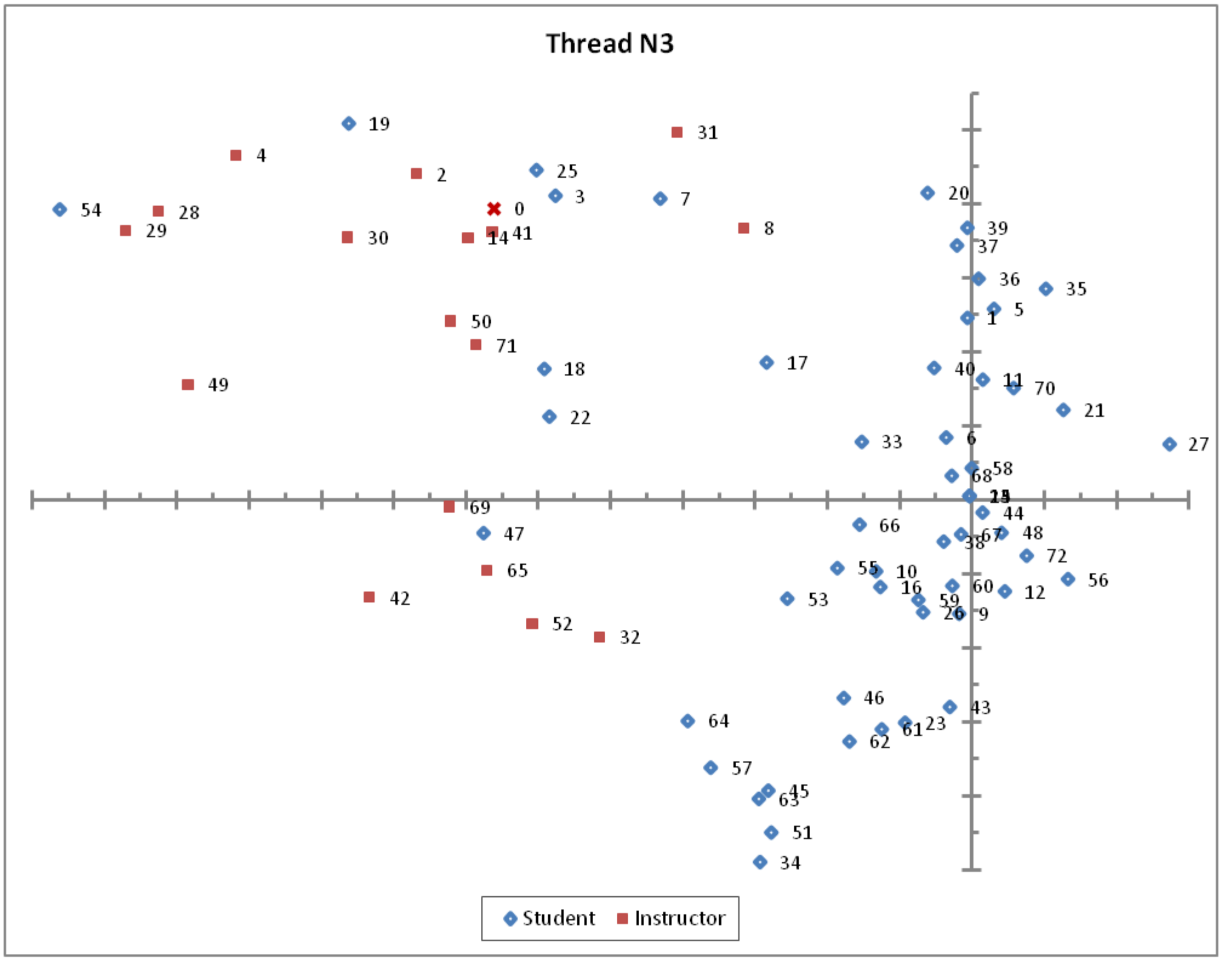

Figure 18. Topic space visualization from discussion thread N3, about the role of natural selection in evolution.

IV. SUMMARY OF FINDINGS

A. Post length, frequency, and timing

For ease of comparison, the quantitative metrics of posting frequency and length for each discussion thread are compiled in **Error! Reference source not found.**, along with a summary of their qualitative characteristics. These metrics reveal that, beyond ensuring that the instructor meets some minimum level of participation, increasing the amount of posting by instructor or students is neither a guarantee nor an indicator of a higher-quality discussion. While students wrote more posts, longer posts, and more total words in threads N2 and N4, many of those posts wandered off-topic rather than delving more deeply into the concepts. In some cases, longer posts came from repeating what students had read elsewhere; while they may have processed the ideas deeply, their posts did not provide clear evidence of this. Nor was post

length by instructor a meaningful indicator of discussion quality. Some of Instructor I's long posts in N4 encouraged more off-topic discussion than biologically-focused discussion. Further, the similarly short posts by Instructors M and P (in L5 and N2, respectively) served different functions in those threads. In the former case, these short posts challenged students to refine their claims and look up new information to add; in the latter case, these short posts only occasionally elicited further elaboration from the students, typically in the form of an example or rarely (twice) a short explanation of the concept. Thus, post length does not capture the value of what an instructor adds to a discussion.

The one parameter among the above that may be associated with discussion quality is instructor's posting frequency, particularly when considered in conjunction with the posts' timing. Thread N2 provides a clear demonstration of the importance of timing, in that Instructor P's early interventions did influence students' subsequent postings, whereas misconceptions emerged and persisted in the absence of guidance in the later part of the discussion. Similarly, misconceptions arose early in thread N4 and were not addressed by Instructor I until after several students had stopped posting. These two examples underscore the importance of monitoring and participating in the discussion both early and consistently throughout its duration, to provide meaningful feedback to all the students' ideas. As thread N3 demonstrates, frequency itself may not matter as much as timing, insofar as Instructor B posted less frequently but more consistently than Instructors P and I, with fewer obvious student misconceptions prevailing. Still, the very high posting frequency of Instructor M in thread L5 took advantage of multiple opportunities to correct, probe, and guide students' thinking and may be partly responsible for the depth of discussion that ensued.

B. Post content

Informative though post frequency and timing may be, they still do not capture the critical role of the content of students' and instructor's posts. The topic modeling and topic space visualizations can provide a valuable window on post content, particularly the relationship between student and instructor posts. Lack of overlap may represent nonspecific acknowledgment or a complete lack of reinforcement (as in thread N2), selective reinforcement (N4, N5), use of different language (N3), or the instructor's attempts to shift the discussion in a particular direction not taken up by the students (N3, and to a lesser extent L5). One caveat is that lack of overlap does not necessarily indicate a lack of communication or understanding between instructor and students. It is possible that the sophisticated language used by Instructor B in thread N3 helped students develop more familiarity with these terms, even if they did not adopt those terms themselves. This cannot be determined without assessing the students' understanding more thoroughly by other means.

At minimum, areas of overlap show where the instructor acknowledged the students' ideas using similar language, although this does not guarantee deep probing of understanding. Closer overlap between student and instructor posts generally indicates a more productive discussion (L5), with both groups of participants talking about the same topics in language that the other understands. However, this overlap needs to occur in regions of desired discussion about important topics to be most valuable. Threads N2, N4, and N5 all show considerable overlap in some parts of the topic space, but unfortunately these did not always correspond with key concepts about natural selection, the intended topic of discussion. (The following section will address techniques for improving the analyses to highlight this more effectively.)

Finally, results from close reading of the discussion threads emphasize the importance of identifying and analyzing the specific behaviors demonstrated by the discussion participants. The patterns observed here suggest that what matters, in terms of instructor intervention, is refining students' knowledge, by probing their understanding, correcting errors and misconceptions, and encouraging more precise explanations, as demonstrated by Instructor M in thread L5. Elaborating on students' knowledge by providing the information directly (N3) appears to have been less effective than offering hints, asking concrete questions, and soliciting it from the students (L5, N4, N5). This could be because the students needed to process the information more deeply by searching for it and explaining its significance themselves. In

addition, acknowledging students' contributions with specific comments about their content (L5, N3, N4, N5) conveyed more information than simply thanking them for posting (N2). These observations suggest that more detailed analyses that incorporate a coding scheme to distinguish these key behaviors may help elucidate patterns in how such facilitation strategies impact discussion outcomes. Combining this human intelligence with the machine intelligence of text mining could yield very powerful analytical techniques for detecting key patterns in online class discussions.

V. INSTRUCTIONAL IMPLICATIONS FOR FACILITATING ONLINE CLASS DISCUSSIONS

Almost all of the instructors in these case studies satisfied and surpassed the expectations enumerated in their faculty review process, further incorporating many strategies highlighted in their training guidelines and the literature reviewed previously. While students and instructors alike were actively engaged in discussing topics related to biology, our analyses applied more stringent standards in evaluating students' demonstration of a nuanced understanding of the mechanisms of natural selection. Holding this as the goal, our results suggest that some of the above facilitation guidelines could perhaps be relaxed and others instituted instead. Rather than seeking to meet specific criteria for word counts, it may be more productive to focus on post timing and frequency instead. Carefully-designed technology can help faculty monitor the discussions and their own behavior by flagging key opportunities for intervention with recommended strategies at critical moments. These may include reining in or redirecting off-topic digressions, addressing prominent or resistant misconceptions, or rephrasing their acknowledgments and comments to more closely mirror students' language. Broader guidelines may be to correct errors, demand precise explanations, ask concrete questions, and elicit specific information from the students, instead of simply providing elaboration and encouraging general connections to the material. In these discussions, promoting more participation, adding new ideas, inviting personal connections, requesting examples, and exploring possible applications were not enough to resolve misconceptions or examine key concepts in depth. Rather, these findings highlight the importance of identifying and probing specific content where students hold divergent or non-normative beliefs, with the goal of helping them to resolve these conflicting ideas and approach normative understanding.

Table 3. Summary of qualitative observations and quantitative metrics on posting frequency and length by participant type for all case studies. (Inst = Instructor, Stud = Student)

Thread	Inst	Instructor Activity	Discussion Characteristics	Discussion quality	Inst/ Stud	Active particip	# Posts			# Words		Post Length by Particip		Active Time (days)		Course grade	
							Total	Avg	SD	Total	Avg	Avg	SD	Avg	SD	Avg	SD
L3	F	No intervention after 1st question	Discussion question isolated in topic space Topics more personal than scientific Whether traits are learned or inherited (not whether learned traits are inherited)	Low	Inst	1	1			78		78.00		0.00			
					Stud	18	55	3.06	2.24	5941	330.06	145.66	82.71	2.35	2.65	91.06	5.50
					Total	19	56										
L5	M	Frequent, continual probing Acknowledged contributions Refined imprecise claims Solicited new information	High student-instructor post overlap Thoughtful reflections Detailed exploration of mechanisms Arguments with supporting sources	High	Inst	1	43			1077		25.05		17.73			
					Stud	17	62	3.65	2.06	5818	342.24	101.98	45.32	6.02	6.54	83.85	15.54
					Total	18	105										
N2	P	Early, brief intervention Nonspecific acknowledgment Requests for examples Minimal questioning, elaborating	Partial student-instructor post overlap Personal anecdotes Confusion over learned vs. inherited traits Creationism suggested as valid alternative	Medium-low	Inst	1	22			764		34.73		3.05			
					Stud	17	115	6.82	3.24	10714	630.24	94.74	42.82	6.16	3.51	88.34	16.04
					Total	18	138										
N3	B	Sustained elaboration Introduced new ideas to consider Thoughtful, sophisticated questions Technical, scientific language	Low student-instructor post overlap Briefly addressed altruism, sexual selection More engaged in general-interest topics Everyday language	Medium-high	Inst	1	18			1373		76.28		6.21			
					Stud	14	55	3.93	1.69	4304	307.43	84.19	24.54	2.75	1.52	N/A	N/A
					Total	15	73										
N4	I	Late intervention Friendly but off-topic elaboration Several posts on peripheral topics Addressed misconceptions, but late	Moderate student-instructor post overlap More definitions than analysis Technology interfering with natl selection Creationism claimed superior to evolution	Medium	Inst	1	14			1925		137.50		7.98			
					Stud	15	89	5.93	2.96	15017	1001.13	178.91	54.18	5.25	3.78	87.07	6.54
					Total	16	103										
N5	I	Lengthy, off-topic elaboration Invited non-biological applications Addressed misconceptions directly Somewhat rhetorical questions	Moderate student-instructor post overlap Natural selection as survival of fittest Applications missed crucial concepts Misconceptions as support for creationism	Medium	Inst	1	21			3026		144.10		4.83			
					Stud	18	120	6.67	2.59	20325	1129.17	172.62	31.21	3.97	2.12	83.92	9.83
					Total	19	141										

VI. CONCERNS AND IMPLICATIONS FOR TEXT MINING CLASS DISCUSSIONS

Much of the prior research that has been done on text mining has investigated sources such as product reviews, which seek to capture consumers' primary concerns, or blogs and discussion forums, which draw visitors based on common interests. These are considerably different from class discussions with mandatory participation requirements, not just in content but also in interaction patterns. As can happen with any class requirement, students may focus on completion rather than comprehension, simply striving to satisfy minimal expectations. Whereas product reviews typically focus on topics about which the respondents have robust knowledge (*i.e.,* their personal experiences and opinions about those experiences), mandatory class discussions expect students to demonstrate knowledge about new and potentially difficult concepts. Whether deliberately or innocently, students may compose responses with the surface appearance of addressing the target ideas, but without truly understanding what they are writing. This decrease in data quality makes it more difficult to rely on text mining to reveal the knowledge and beliefs embedded within the discussions.

Further, as previously described in the Methods section, the text mining analyses here were conducted on collections of discussion threads around similar questions, without initially training the system on a preselected body of text representing the "ideal" concepts, language, or discussion patterns. The motivation for this was to explore patterns that naturally emerged from the topic models of the discussions, without biasing them toward or against prior expectations. What the results revealed was that the amount of variability in what students chose to discuss overshadowed the central concepts, which were then difficult to detect in the topic models and visualizations produced. Even though our analyses revealed a valuable contrast between a very productive discussion (L5) and several less-focused discussions, the differences in the patterns that emerged were masked by the variability in discussion content. The PCA projections selected the dimensions of greatest variability for visualizing the discussions, rather than the dimensions of greatest conceptual importance. Training the system on a sample of text containing the desired concepts would thus enable the analyses to highlight when the discussion addressed these topics and to better differentiate that from off-topic or less-relevant conversation. Such a sample could be drawn from course texts and other authoritative sources on the desired topic. This could then capture whether students discuss key concepts such as inheritance and reproduction, perhaps signaling instructors so that they can watch for and guide the discussion toward these concepts.

In addition, since the learning goals are already known in advance, the model can incorporate more domain- and task-specific information on the desired outcomes up front. By training the system on sample student text that has already been labeled (*i.e.,* graded), the system can then explicitly seek out evidence of having met those goals (cf. 36). This could include exemplars of good student work as well as common misconceptions that may emerge and can be addressed in the discussions. Mapping the desired concept on one axis and a common misconception on another axis could help elucidate the relative strength of those conceptions as the discussion develops.

Even with this additional training, text mining alone may not be sufficient for determining whether students are using key terms correctly, much less whether they understand the concepts. This is especially difficult in science classes, where the terms need to appear in particular sequences and relationships when explaining causal mechanisms. An appropriate role for text mining here may be to flag important terms and concepts for expert human judges (*i.e.,* instructors) to evaluate if the students are using them correctly and to intervene as needed. While more sophisticated models than LSA (e.g., pLSA, LDA, author-topic

models) and other techniques beyond PCA (e.g., LLE, IsoMap, network models) should improve the capabilities of the system, there is no reason to expect them to completely address these concerns.

Another significant challenge is in assessing whether students are applying or transferring their knowledge from class to new contexts. Educators often encourage students to relate what they are learning to their personal experiences or to generate a novel example. Without training the model on these new examples, it has no basis for determining whether students applied their knowledge correctly. Achieving this using text mining would require drawing upon a much richer domain model to evaluate the accuracy and appropriateness of students' attempts to go beyond the original text.

In spite of the above caveats, text mining on mandatory class discussion forums can still offer worthwhile insights, particularly in regard to capturing the relationships between the content of students' and instructor's posts. As shown here, closer overlap indicates where students and instructor are discussing the same concepts and is generally preferable, as long as the overlap occurs around desirable areas of the topic space. Identifying these areas can be facilitated by training the topic model in advance on target concepts which students ought to discuss, and perhaps also on common misconceptions likely to need further exploration or remediation. Such an approach would also enable tracking the ebb and flow of individual topics to reveal crucial intervention points; this visualization technique was explored but not discussed in this report due to the challenges of interpreting the particular clusters of topics that emerged from the (untrained) topic model.

VII. CONCLUSIONS

These results demonstrate the power of analyzing classroom discussions by applying LSA, a valuable topic modeling technique for its capacity to quickly identify key patterns from among a large quantity of text data. Such a technique can accelerate the analysis process for researchers and faculty alike by surveying many years of archived as well as live discussion data, aggregating common patterns, and flagging major concerns for human readers to examine and address in more depth. Among the many potential applications for this research, the most immediate follow-up would be working with instructors to determine which information from these analysis techniques they find most useful for improving their facilitation strategies. While these analyses can help researchers or faculty trainers understand class discussions retrospectively, faculty may have slightly different experiences when facilitating discussions in real time. Both researchers and developers need to collaborate with faculty and faculty trainers to ensure the relevance of tools based on the research reported here.

This work has a number of additional future directions and longer-term applications. These techniques could be applied to provide feedback to students on the nature of the discussions, perhaps improving reflective learning and metacognition. By analyzing large quantities of past discussions, implementing a "find similar posts" or "find similar discussions" feature could allow instructors or students to locate previous posts or discussion threads about similar topics. These techniques could be used to generate topical summaries of large amounts of discussion forum data quickly and easily.

Other possible extensions of these visualization tools may be to combine the machine intelligence with human intelligence. Ideally, good instruction should go beyond monitoring compliance with requirements and focus on supporting the continued development of understanding. Capitalizing on students' knowledge and instructors' expertise to recognize and label key discussion behaviors or facilitation strategies may augment the text mining (cf. 36). Human users, whether students or instructors, could provide some preliminary categorization, coding, or tagging of their own posts along a specified dimension of interest. These could focus on domain-general behaviors relating to collaborative discourse processes (questions, answers, comments), higher-order thinking and argumentation skills (claim,

evidence, justification), or facilitation strategies (acknowledge, refine, elaborate, extend). Filtering along some dimension can then enable participants and observers to become more aware of the activities taking place in the discussion, thereby encouraging greater metacognitive self-regulation and enabling more effective intervention.

Accounting for other discussion characteristics that go beyond the text enables using a broader range of analytical techniques that may further enhance the value of the text mining. Participating in collaborative discourse should add more value to the instructional experience than simply submitting assignments individually. Incorporating social network analysis (*e.g.*, 20 22), may enable closer examination of how participants' roles interact with the collective construction of understanding in the discussion. Such approaches may enrich the analyses with a more nuanced picture of how the participants are interacting and what they are learning from the discussions, since the text alone is a limited source of information on these phenomena.

Importantly, the text mining techniques presented here do not speak entirely for themselves. Without combining their results with closer analyses of the discussion content and interactions, they cannot tell us that one instructor was effective at facilitating a discussion while another was ineffective. What they can do, however, is draw attention to latent patterns in discussion data, patterns that might otherwise go unnoticed, and make those patterns readily visible and interpretable. This approach means taking advantage of computers' computational strengths, in terms of analyzing vast quantities of data with relative speed and ease, while simultaneously taking advantage of humans' cognitive strengths, in terms of interpreting and making meaning from the results of those analyses. Thus, this research aims not to replace faculty or faculty trainers, but rather to provide computational tools that both support their current activities and enable new activities.

VIII. ABOUT THE AUTHORS

Norma C. Ming is Senior Research Scientist at the Nexus Research and Policy Center. Her current research interests encompass instructional design in math and science education, how people learn from comparisons, and professional development of teachers and faculty. She is particularly interested in merging research with cutting-edge technology to create innovative, assessment-driven teaching and learning environments that combine machine intelligence with collaborative human intelligence to advance both insights and outcomes. Prior to joining Nexus, Dr. Ming was a lecturer in Education in Math, Science, and Technology at UC Berkeley's Graduate School of Education, where she is now a visiting scholar. She earned an A.B. with honors in chemistry at Harvard University and a Ph.D. in cognitive psychology in the Program for Interdisciplinary Educational Research at Carnegie Mellon University.

Eric P. S. Baumer is a postdoctoral associate in the Information Science Department at Cornell University. His research focuses in large part on how computational analyses can be incorporated into tools and systems that promote critical thinking among human users. He employs an interdisciplinary perspective, working in such areas as human-computer interaction, natural language processing, and new media & society, and has published in conference proceedings and journal papers across a broad range of disciplines. His current research focuses on using analysis of online political content to support political discussion and deliberation. He holds a B.S. magna cum laude in Computer Science with a Minor in Music from the University of Central Florida, and M.S. and Ph.D. degrees in Information and Computer Sciences from the University of California, Irvine.

IX. ACKNOWLEDGMENTS

We greatly appreciate the expertise of Cristen Torrey in sharing valuable resources and guidance in the early stages of the research, processing raw forum data into readable discussion threads, and offering valuable insights for analyzing the discussions. We are likewise grateful to David Holstius and Vivienne Ming for their assistance in preprocessing the text so that we could parse and read the data more readily. We also thank Shelley DeFord and the members of the Torrey-Ming research group at UC Berkeley's Graduate School of Education for their perceptive comments in examining numerous discussion threads whose analyses lay the foundation for this research. Finally, we appreciate the constructive comments and feedback provided by anonymous reviewers.

X. REFERENCES

1. **Asterhan, C.S.C. & Schwarz, B.B.** "Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog," *Cognitive Science*, vol. 33, pp. 374-400, 2009.
2. **Baker, M.J.; deVries, E.; Lund, K. & Quignard, M.** "Computer-mediated epistemic interactions for co-constructing scientific notions: Lessons learned from a five-year research program," in *European Perspectives on CSCL* (P. Dillenbourg, A. Eurelings, & K. Hakkarainnen, Eds.), 2001, pp. 89-96.
3. **McLaren, B.M.; Scheuer, O. & Miksatko, J.** "Supporting collaborative learning and e-discussions using artificial intelligence techniques," *International Journal of Artificial Intelligence in Education*, vol. 20, no. 1, pp. 1-46, 2010.
4. **Scheuer, O.; Loll, F.; Pinkwart, N. & McLaren, B.M.** "Computer-supported argumentation: A review of the state of the art," *International Journal of Computer-Supported Collaborative Learning*, vol. 5, no. 1, pp. 43-102, 2010.
5. **Okada, T. & Simon, H.A.** "Collaborative discovery in a scientific domain," *Cognitive Science*, vol. 21, no. 2, pp. 109-146, 1997.
6. **Webb, N.M. & Mastergeorge, A.M.** "The development of students' helping behavior and learning in peer-directed small groups," *Cognition and Instruction*, vol. 21, no. 4, pp. 361-428, 2003.
7. **Sandoval, W.A. & Millwood, K.A.** "The quality of students' use of evidence in written scientific explanations," *Cognition and Instruction*, vol. 23, no. 1, pp. 23-55, 2005.
8. **Schwartz, R. & Lederman, N. "**What scientists say: Scientists' views of nature of science and relation to science context," *International Journal of Science Education*, vol. 30, no. 6, pp. 727-771, 2008.
9. **Webb, N.M.** "The teacher's role in promoting collaborative dialogue in the classroom," *British Journal of Educational Psychology*, vol. 79, pp. 1-28, 2009.
10. **Rovai, A.P.** "Facilitating online discussions effectively," *Internet and Higher Education*, vol. 10, pp. 77-88, 2007.
11. **Collison, G.; Elbaum, B.; Haavind, S. & Tinker, R.** *Facilitating online learning: Effective strategies for moderators*. Madison WI: Atwood Publishing, 2000.
12. **Hammond, M.** "A review of recent papers on online discussion in teaching and learning in higher education," *Journal of Asynchronous Learning Networks*, vol. 9, no. 3, pp. 9-23, 2005.
13. **Anderson, T.; Rourke, L.; Garrison, R. & Archer, W.** "Assessing teacher presence in a computer conferencing context," *Journal of Asynchronous Learning Networks*, vol. 5, no. 2, pp. 1-17, 2001.
14. **Gilbert, P.K. & Dabbagh, N.** "How to structure online discussions for meaningful discourse: a case study," *British Journal of Educational Psychology*, vol. 36, no. 1, pp. 5-18, 2005.
15. **Oliver, M. & Shaw, G.P.** "Asynchronous discussion in support of medical education," *Journal of Asynchronous Learning Networks*, vol. 7, no. 1, pp. 56-67, 2003.
16. **Chi, M.T.H.** "Quantifying qualitative analyses of verbal data: A practical guide," *Journal of the Learning Sciences*, vol. 6, no. 3, pp. 271-315, 1997.

17. **Ericsson, K.A. & Simon, H.A.** *Protocol analysis: Verbal reports as data*. Cambridge MA: MIT Press, 1984.
18. **Hicks, D.** “Discourse, learning, and teaching,” *Review of Research in Education*, vol. 21, pp. 49-95, 1995 - 1996.
19. **Faridani, S.; Bitton, E.; Ryokai, K. & Goldberg, K.** “Opinion Space: A Scalable Tool for Browsing Online Comments.” In *Proceedings of the ACM Conference on Human Factors in Computer Systems*. Atlanta, GA: ACM Press, 2010, pp. 1175-1184. doi:http://doi.acm.org/10.1145/1753326.1753502
20. **Fisher, D.; Smith, M. & Welser, H.T.** “You Are Who You Talk To: Detecting Roles in Usenet Newsgroups.” In *Hawai'i International Conference on System Sciences (HICSS)*. Kauai, HI: IEEE Computer Society, 2006. doi:10.1109/HICSS.2006.536
21. **Viégas, F.B. & Smith, M.** “Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals in Conversational Cyberspaces.” In *Hawai'i International Conference on System Sciences (HICSS)*. Big Island, HI: IEEE Computer Society, 2004.
22. **Welser, H.T.; Gleave, E.; Fisher, D. & Smith, M.** "Visualizing the signatures of social roles in online discussion groups," *Journal of Social Structure*, vol. 8, no. 2, 2007.
23. **Ai, H.; Sionti, M.; Wang, Y.C. & Rose, C.** "Finding Transactive Contributions in Whole Group Classroom Discussions," in *Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers* (K. Gomez, L. Lyons, & J. Radinsky, Eds.). Chicago IL: International Society of the Learning Sciences, 2010.
24. **Ming, N.C.** “Strategies for Facilitating Online Discussion.” San Francisco, CA: Nexus Research and Policy Center, 2010.
25. Latent semantic analysis is sometimes also referred to as latent semantic indexing (LSI), due to its applications in the domain of information retrieval. This report uses the term latent semantic analysis and the abbreviation LSA throughout.
26. **Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K. & Harshman, R.** "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391-407, 1990.
27. **Landauer, T.K.; Foltz, P.W. & Laham, D.** "Introduction to Latent Semantic Analysis," *Discourse Processes*, vol. 25, pp. 259-284, 1998.
28. **Blei, D. & Lafferty, J.** “Topic Models.” In *Text Mining: Theory and Application* (A. Srivastava & M. Sahami, Eds.). Taylor and Francis, 2009. Retrieved from http://www.cs.princeton.edu/~blei/papers/BleiLafferty2009.pdf
29. **Steyvers, M. & Griffiths, T.** "Probabilistic topic models," in *Latent Semantic Analysis: A Road to Meaning* (T. Landauer, D. McNamara, S. Dennis, & W. Kintsch, Eds.). Hillsdale, NJ: Lawrence Erlbaum, 2007.
30. **Feldman, R. & Sanger, J.** “Text Mining Preprocessing Techniques,” in *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, pp. 57-63. Cambridge: Cambridge University Press, 2007.
31. **Nallapati, R. & Cohen, W.** “Link-PLSA-LDA: A New Unsupervised Model for Topics and Influence in Blogs,” in *International Conference on Weblogs and Social Media*, pp. 84-92. Seattle, WA, 2008.
32. **Yano, T.; Cohen, W.W. & Smith, N.A.** “Predicting Response to Political Blog Posts with Topic Models,” in *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pp. 477-485. Boulder, CO, 2009.
33. **Blei, D.; Ng, A. & Jordan, M.** "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 1, pp. 993-1022, 2003.
34. **Hofmann, T.** “Probabilistic latent semantic indexing.” In *Proceedings of the 22nd annual international ACM SIGIR conference on research and development in information retrieval*. Berkeley, California: ACM, 1999, pp. 50-57. doi:10.1145/312624.312649.

35. **Steyvers, M.; Smyth, P.; Rosen-Zvi, M. & Griffiths, T.** “Probabilistic author-topic models for information discovery.” In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Seattle, WA: ACM, 2004, pp. 306-315. doi:10.1145/1014052.1014087.
36. **Foltz, P.W.; Laham, D. & Landauer, T.K.** "Automated essay scoring: Applications to educational technology," in *World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA)*. Chesapeake, VA: AACE, 1999, pp. 939-944.